FIMFA

Intégration, Probabilités et Processus Aléatoires

Jean-François Le Gall

Septembre 2006

Département Mathématiques et Applications
Ecole normale supérieure de Paris

Sommaire

Partie I

Intégration

Chapitre 1

Espaces mesurés

L'idée de départ de la théorie de la mesure est d'assigner un nombre réel positif (la mesure de ce sous-ensemble) à chaque sous-ensemble d'un ensemble donné, de manière à satisfaire certaines propriétés naturelles d'additivité (la mesure d'une réunion disjointe doit être la somme des mesures). Pour des raisons profondes, il n'est pas possible en général de définir la mesure de n'importe quel sous-ensemble, et on doit se restreindre à une certaine classe (tribu) de sous-ensembles, appelés les sous-ensembles mesurables : un ensemble muni d'une tribu est appelé espace mesurable. Ce chapitre introduit les notions fondamentales de tribu (= famille des ensembles mesurables), de mesure sur un espace mesurable, et de fonctions mesurables, qui sont les fonctions dont on saura plus tard définir l'intégrale. Le dernier paragraphe énonce une forme du lemme de classe monotone, qui joue un rôle très important à la fois en théorie de la mesure et en théorie des probabilités.

1.1 Ensembles mesurables

Définition 1.1.1 *Soit E un ensemble quelconque. Une tribu (ou σ-algèbre) sur E est une famille $\mathcal{A}$ de parties de E telle que:*

(i) $E \in \mathcal{A}$;

(ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$;

(iii) Si $A_n \in \mathcal{A}$ pour tout $n \in \mathbb{N}$, on a aussi $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Les éléments de $\mathcal{A}$ sont appelés parties mesurables, ou parfois $\mathcal{A}$-mesurables s'il y a ambiguïté. On dit que $(E, \mathcal{A})$ est un espace mesurable.

Enonçons quelques conséquences de la définition :

(1) $\varnothing \in \mathcal{A}$

(2) Si $A_n \in \mathcal{A}$ pour tout $n \in \mathbb{N}$, on a aussi $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

(3) Puisqu'on peut toujours prendre $A_n = \varnothing$ pour n assez grand, la propriété (iii) entraîne que $\mathcal{A}$ est stable par réunions finies (et de même par intersection finies).

Exemples.

- $\mathcal{A} = \mathcal{P}(E)$;
- $\mathcal{A} = \{\varnothing, E\}$ est la tribu triviale ;
- l'ensemble des parties de E qui sont (au plus) dénombrables ou dont le complémentaire est (au plus) dénombrable forme une tribu sur E.

Pour donner des exemples plus intéressants, on remarque qu'une intersection quelconque de tribus est encore une tribu. Ceci conduit à la définition suivante.

Définition 1.1.2 *Soit $\mathcal{C}$ un sous-ensemble de $\mathcal{P}(E)$. Il existe alors une plus petite tribu sur E qui contienne $\mathcal{C}$. Cette tribu notée $\sigma(\mathcal{C})$ peut être définie par*

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{A} \text{ tribu}, \mathcal{C} \subset \mathcal{A}} \mathcal{A}\,.$$

$\sigma(\mathcal{C})$ est appelée la tribu engendrée par $\mathcal{C}$.

Tribu borélienne. Pour donner un premier exemple de l'intérêt de la notion de tribu engendrée, considérons le cas où E est un espace topologique.

Définition 1.1.3 *Supposons que E est un espace topologique, et soit $\mathcal{O}$ la classe des ouverts de E. La tribu $\sigma(\mathcal{O})$ est appelée tribu borélienne et notée $\mathcal{B}(E)$.*

La tribu borélienne est donc la plus petite tribu qui contienne tous les ouverts de E. Les éléments de $\mathcal{B}(E)$ sont appelés boréliens de E.

Dans la suite, à chaque fois que l'on considérera un espace topologique, par exemple $\mathbb{R}$ ou $\mathbb{R}^d$, on supposera sauf indication du contraire qu'il est muni de sa tribu borélienne.

Exercice. Vérifier que la tribu $\mathcal{B}(\mathbb{R})$ est aussi engendrée par les intervalles $]a, b[$, $a, b \in \mathbb{R}$, $a < b$, ou par les intervalles $]-\infty, a[$, $a \in \mathbb{R}$, ou encore les intervalles $]-\infty, a[$, $a \in \mathbb{Q}$ (on peut aussi remplacer intervalles ouverts par intervalles fermés).

Tribu-produit. Un deuxième exemple important de la notion de tribu engendrée est la tribu-produit.

Définition 1.1.4 *Soient $(E_1, \mathcal{A}_1)$ et $(E_2, \mathcal{A}_2)$ deux espaces mesurables. La tribu-produit est la tribu sur $E_1 \times E_2$ définie par*

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(A_1 \times A_2\,;\, A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}.$$

Exercice. Vérifier que

$$\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}).$$

1.2 Mesures positives

Soit $(E, \mathcal{A})$ un espace mesurable.

Définition 1.2.1 *Une mesure positive sur $(E, \mathcal{A})$ est une application $\mu : \mathcal{A} \longrightarrow [0, \infty]$ qui vérifie les propriétés suivantes:*

(i) $\mu(\varnothing) = 0$;

(ii) Pour toute famille $(A_n)_{n\in\mathbb{N}}$ de parties mesurables disjointes,

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

Remarquons qu'il est important d'autoriser la valeur $+\infty$. La propriété (ii) est appelée σ-additivité. Elle contient évidemment le cas particulier où les A_n sont vides à partir d'un certain rang, ce qui donne la propriété d'additivité finie.

Propriétés.

(1) Si $A \subset B$, $\mu(A) \leq \mu(B)$ et si de plus $\mu(A) < \infty$,

$$\mu(B\backslash A) = \mu(B) - \mu(A)\,;$$

(2) Si $A, B \in \mathcal{A}$,

$$\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)\,;$$

(3) Si $A_n \in \mathcal{A}$ et $A_n \subset A_{n+1}$,

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \uparrow \mu(A_n)\,;$$

(4) Si $B_n \in \mathcal{A}$ et $B_{n+1} \subset B_n$, et si $\mu(B_0) < \infty$,

$$\mu(\bigcap_{n\in\mathbb{N}} B_n) = \lim_{n\to\infty} \downarrow \mu(B_n)\,;$$

(5) Si $A_n \in \mathcal{A}$,

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} \mu(A_n).$$

Démontrons seulement (3),(4) et (5). Pour (3), on pose $C_0 = A_0$ et pour tout $n \geq 1$,

$$C_n = A_n \backslash A_{n-1}$$

de sorte que $\cup A_n = \cup C_n$. Puisque les C_n sont disjoints,

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) = \mu(\bigcup_{n\in\mathbb{N}} C_n) = \sum_{n\in\mathbb{N}} \mu(C_n) = \lim_{N\to\infty} \uparrow \sum_{n=0}^{N} \mu(C_n) = \lim_{N\to\infty} \uparrow \mu(A_N).$$

Pour (4), on pose $A_n = B_0 \backslash B_n$ pour tout n, de sorte que la suite (A_n) est croissante. Alors

$$\mu(B_0) - \mu(\bigcap_{n\in\mathbb{N}} B_n) = \mu(B_0 \backslash \bigcap_{n\in\mathbb{N}} B_n) = \mu(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \uparrow \mu(A_n) = \lim_{n\to\infty} \uparrow (\mu(B_0) - \mu(B_n)).$$

La condition $\mu(B_0) < \infty$ est utilisée notamment pour écrire $\mu(A_n) = \mu(B_0) - \mu(B_n)$.

Enfin, pour (5), on pose $C_0 = A_0$ puis pour tout $n \geq 1$,

$$C_n = A_n \backslash \bigcup_{k=0}^{n-1} A_k.$$

Les ensembles C_n sont disjoints et donc

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) = \mu(\bigcup_{n\in\mathbb{N}} C_n) = \sum_{n\in\mathbb{N}} \mu(C_n) \leq \sum_{n\in\mathbb{N}} \mu(A_n).$$

Exemples.

(1) Si $E = \mathbb{N}$, et $\mathcal{A} = \mathcal{P}(\mathbb{N})$, la mesure de comptage est définie par

$$\mu(A) = \mathrm{Card}(A).$$

(On peut définir plus généralement la mesure de comptage sur $(E, \mathcal{P}(E))$ lorsque E est quelconque.) Cet exemple permet de voir que la condition $\mu(B_0) < \infty$ est nécessaire dans la propriété (4) ci-dessus : en prenant

$$B_n = \{n, n+1, n+2, \ldots\}$$

on a $\mu(B_n) = \infty$ alors que $\cap B_n = \varnothing$ et donc $\mu(\cap B_n) = 0$.

(2) Soit $(E, \mathcal{A})$ quelconque et soit $x \in E$. La mesure δ_x définie par

$$\delta_x(A) = 1_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

est appelée mesure de Dirac au point x. Plus généralement, si x_n, $n \in \mathbb{N}$ sont des points de E et $\alpha_n \in [0, \infty]$ on peut considérer la mesure $\sum \alpha_n \delta_{x_n}$ définie par

$$(\sum \alpha_n \delta_{x_n})(A) = \sum \alpha_n \delta_{x_n}(A) = \sum \alpha_n 1_A(x_n).$$

(3) Mesure de Lebesgue. Il existe une unique mesure positive sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, notée λ, telle que pour tout intervalle ouvert $]a, b[$ de $\mathbb{R}$ on ait $\lambda(]a, b[) = b - a$. L'existence et l'unicité de cette mesure seront établies plus loin.

Définitions.

- μ est dite finie si $\mu(E) < \infty$ (la quantité $\mu(E)$ est la masse totale de μ).
- μ est une mesure de probabilité si $\mu(E) = 1$.
- μ est dite σ-finie s'il existe une suite croissante de parties mesurables E_n telles que $E = \bigcup_{n\in\mathbb{N}} E_n$ et $\mu(E_n) < \infty$ pour tout n.
- $x \in E$ est un atome de μ si $\mu(\{x\}) > 0$ (on suppose que $\{x\} \in \mathcal{A}$).
- La mesure μ est dite diffuse si elle n'a pas d'atomes.

1.3 Fonctions mesurables

Définition 1.3.1 *Soient* $(E, \mathcal{A})$ *et* $(F, \mathcal{B})$ *deux espaces mesurables. Une application* $f : E \longrightarrow F$ *est dite mesurable si*

$$\forall B \in \mathcal{B} \ , \ f^{-1}(B) \in \mathcal{A}.$$

Lorsque E *et* F *sont des espaces topologiques munis de leurs tribus boréliennes, on dit aussi que* f *est borélienne.*

Proposition 1.3.1 *La composition de deux applications mesurables est encore mesurable.*

C'est immédiat en écrivant $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$.

Proposition 1.3.2 *Pour que* f *soit mesurable, il suffit qu'il existe une sous-classe* $\mathcal{C}$ *de* $\mathcal{B}$ *telle que* $\sigma(\mathcal{C}) = \mathcal{B}$ *et telle que la propriété* $f^{-1}(B) \in \mathcal{A}$ *soit vraie pour tout* $B \in \mathcal{C}$.

Preuve. Soit

$$\mathcal{G} = \{B \in \mathcal{B} : f^{-1}(B) \in \mathcal{A}\}.$$

Alors il est facile de vérifier que $\mathcal{G}$ est une tribu. Par hypothèse $\mathcal{C} \subset \mathcal{G}$. Il en découle que $\mathcal{G}$ contient $\sigma(\mathcal{C}) = \mathcal{B}$, d'où le résultat recherché. □

Exemples. (1) Dans le cas où $(F, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(R))$, il suffit pour montrer que f est mesurable d'établir que les ensembles $f^{-1}(]a, b[)$, ou même les $f^{-1}(]-\infty, a[)$ sont mesurables.

(2) Dans le cas où E et F sont des espaces topologiques munis de leurs tribus boréliennes, toute application continue est aussi mesurable (prendre pour $\mathcal{C}$ la classe des ouverts de F).

Opérations sur les fonctions mesurables.

Lemme 1.3.3 *Soient* $f_1 : (E, \mathcal{A}) \longrightarrow (F_1, \mathcal{B}_1)$ *et* $f_2 : (E, \mathcal{A}) \longrightarrow (F_2, \mathcal{B}_2)$ *deux applications mesurables. Alors l'application produit* $f : (E, \mathcal{A}) \longrightarrow (F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ *définie par* $f(x) = (f_1(x), f_2(x))$ *est aussi mesurable.*

Preuve. On applique la dernière proposition en prenant

$$\mathcal{C} = \{B_1 \times B_2 \,;\, B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2\}.$$

Puisque $f^{-1}(B_1 \times B_2) = f_1^{-1}(B_1) \cap f_2^{-1}(B_2) \in \mathcal{A}$ on obtient immédiatement le résultat. □

Remarque. La réciproque de la proposition (si f est mesurable, f_1 et f_2 le sont aussi) est vraie et aussi facile.

Corollaire 1.3.4 *Si* $f, g : (E, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{B}(R))$ *sont mesurables, alors les fonctions* $f + g$, fg, $\inf(f, g)$, $f^+ = \sup(f, 0)$, $f^- = \sup(-f, 0)$ *sont mesurables.*

La démonstration est facile : par exemple $f+g$ est la composée des deux applications $x \longrightarrow (f(x), g(x))$ et $(a,b) \longrightarrow a+b$ qui sont mesurables, la seconde parce que continue.

Rappelons que si (a_n) est une suite d'éléments de $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, on définit

$$\limsup a_n = \lim_{n\to\infty} \downarrow \left(\sup_{k\geq n} a_k\right) \ , \ \liminf a_n = \lim_{n\to\infty} \uparrow \left(\inf_{k\geq n} a_k\right),$$

les limites existant dans $\bar{\mathbb{R}}$. Alors, $\limsup a_n$ et $\liminf a_n$ sont respectivement la plus grande et la plus petite valeur d'adhérence de la suite (a_n).

Proposition 1.3.5 *Si f_n est une suite de fonctions mesurables de E dans $\bar{\mathbb{R}}$, alors*

$$\sup_n f_n \ , \ \inf_n f_n \ , \ \limsup f_n \ , \ \liminf f_n$$

sont aussi mesurables. En particulier si la suite f_n converge simplement, sa limite $\lim f_n$ *est mesurable. En général, l'ensemble* $\{x \in E : \lim f_n(x) \text{ existe}\}$ *est mesurable.*

Preuve. Soit $f(x) = \inf f_n(x)$. Il suffit de montrer que pour tout $a \in \mathbb{R}$, $f^{-1}([-\infty, a[) \in \mathcal{A}$. Or

$$f^{-1}([-\infty, a[) = \{x : \inf f_n(x) < a\} = \bigcup_n \{x : f_n(x) < a\}$$

d'où le résultat. On traite de même le cas de $\sup f_n$.

Il en découle que

$$\liminf f_n = \sup_{n\geq 0} \left(\inf_{k\geq n} f_k\right)$$

est mesurable.

Pour la dernière assertion, on écrit

$$\{x \in E : \lim f_n(x) \text{ existe}\} = \{x \in E : \liminf f_n(x) = \limsup f_n(x)\} = G^{-1}(\Delta)$$

si G est l'application mesurable $G(x) = (\liminf f_n(x), \limsup f_n(x))$ et Δ désigne la diagonale de $\bar{\mathbb{R}}^2$, qui est mesurable parce que fermée. □

Notion de mesure-image.

Définition 1.3.2 *Soit $f : (E, \mathcal{A}) \longrightarrow (F, \mathcal{B})$ une application mesurable, et soit μ une mesure positive sur $(E, \mathcal{A})$. La mesure-image de μ par f, notée $f(\mu)$ est la mesure positive sur $(F, \mathcal{B})$ définie par*

$$f(\mu)(B) = \mu(f^{-1}(B)).$$

Il est facile de voir que la dernière formule définit bien une mesure sur $(F, \mathcal{B})$. Les mesures μ et $f(\mu)$ ont même masse totale, mais il peut arriver que μ soit σ-finie sans que $f(\mu)$ le soit.

1.4 Classe monotone

Définition 1.4.1 *Un sous-ensemble $\mathcal{M}$ de $\mathcal{P}(E)$ est appelé classe monotone si*

(i) $E \in \mathcal{M}$;

(ii) Si $A, B \in \mathcal{M}$ et $A \subset B$, alors $B \backslash A \in \mathcal{M}$;

(iii) Si $A_n \in \mathcal{M}$ et $A_n \subset A_{n+1}$, alors $\bigcup_n A_n \in \mathcal{M}$.

Toute tribu est aussi une classe monotone. Comme dans le cas des tribus, on voit immédiatement que toute intersection de classes monotones est encore une classe monotone. Si $\mathcal{C}$ est une partie quelconque de $\mathcal{P}(E)$, on peut donc définir la classe monotone engendrée par $\mathcal{C}$, notée $\mathcal{M}(\mathcal{C})$, en posant

$$\mathcal{M}(\mathcal{C}) = \bigcap_{\mathcal{M} \text{ classe monotone}, \ \mathcal{C} \subset \mathcal{M}} \mathcal{M} \, .$$

Théorème 1.4.1 (Lemme de classe monotone) *Si $\mathcal{C} \subset \mathcal{P}(E)$ est stable par intersections finies, alors $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.*

Preuve. Puisque toute tribu est une classe monotone, il est clair qu'on a $\mathcal{M}(\mathcal{C}) \subset \sigma(\mathcal{C})$. Pour établir l'inclusion inverse, il suffit de montrer que $\mathcal{M}(\mathcal{C})$ est une tribu. Or une classe monotone est une tribu si et seulement si elle est stable par intersections finies (en effet, par passage au complémentaire, elle sera alors stable par réunion finies, puis par passage à la limite croissant par réunion dénombrable). Montrons donc que $\mathcal{M}(\mathcal{C})$ est stable par intersections finies.

Soit $A \in \mathcal{C}$ fixé. Posons

$$\mathcal{M}_1 = \{B \in \mathcal{M}(\mathcal{C}) : A \cap B \in \mathcal{M}(\mathcal{C})\}.$$

Puisque $\mathcal{C}$ est stable par intersections finies, il est clair que $\mathcal{C} \subset \mathcal{M}_1$. Vérifions ensuite que $\mathcal{M}_1$ est une classe monotone:

- $E \in \mathcal{M}_1$ est immédiat.
- Si $B, B' \in \mathcal{M}_1$ et $B \subset B'$, on a $A \cap (B' \backslash B) = (A \cap B') \backslash (A \cap B) \in \mathcal{M}(\mathcal{C})$ et donc $B' \backslash B \in \mathcal{M}_1$.
- Si $B_n \in \mathcal{M}_1$ pour tout n et la suite B_n croît, on a $A \cap (\cup B_n) = \cup (A \cap B_n) \in \mathcal{M}(\mathcal{C})$ et donc $\cup B_n \in \mathcal{M}_1$.

Puisque $\mathcal{M}_1$ est une classe monotone qui contient $\mathcal{C}$, $\mathcal{M}_1$ contient aussi $\mathcal{M}(\mathcal{C})$. On a donc montré

$$\forall A \in \mathcal{C}, \ \forall B \in \mathcal{M}(\mathcal{C}), \ A \cap B \in \mathcal{M}(\mathcal{C}).$$

Ce n'est pas encore le résultat recherché, mais on peut appliquer la même idée une seconde fois. Précisément, on fixe maintenant $B \in \mathcal{M}(\mathcal{C})$, et on pose

$$\mathcal{M}_2 = \{A \in \mathcal{M}(\mathcal{C}) : A \cap B \in \mathcal{M}(\mathcal{C})\}.$$

D'après la première étape de la preuve, $\mathcal{C} \subset \mathcal{M}_2$. En reprenant exactement les mêmes arguments que dans la première étape, on montre que $\mathcal{M}_2$ est une classe monotone. Il en découle que $\mathcal{M}(\mathcal{C}) \subset \mathcal{M}_2$, ce qui montre bien que $\mathcal{M}(\mathcal{C})$ est stable par intersections finies et termine la preuve. □

Corollaire 1.4.2 *Soient μ et ν deux mesures sur $(E, \mathcal{A})$. Supposons qu'il existe une classe $\mathcal{C} \subset \mathcal{A}$ stable par intersections finies, telle que $\sigma(\mathcal{C}) = \mathcal{A}$ et $\mu(A) = \nu(A)$ pour tout $A \in \mathcal{C}$.*

(1) Si $\mu(E) = \nu(E) < \infty$, on a $\mu = \nu$.

(2) S'il existe une suite croissante de parties $E_n \in \mathcal{C}$ telles que $E = \cup E_n$ et $\mu(E_n) = \nu(E_n) < \infty$, on a $\mu = \nu$.

Preuve. (1) Soit $\mathcal{G} = \{A \in \mathcal{A} : \mu(A) = \nu(A)\}$. Par hypothèse, $\mathcal{C} \subset \mathcal{G}$. Par ailleurs, on vérifie aisément que $\mathcal{G}$ est une classe monotone : par exemple, si $A, B \in \mathcal{G}$ et $A \subset B$, on a $\mu(B \backslash A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \backslash A)$, et donc $B \backslash A \in E$ (noter qu'on utilise ici le fait que μ et ν sont finies).

On conclut que $\mathcal{G}$ contient $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C}) = \mathcal{A}$ (la première égalité d'après le théorème de classe monotone, la seconde par hypothèse). Donc $\mathcal{G} = \mathcal{A}$, c'est-à-dire $\mu = \nu$.

(2) Notons, pour tout n, μ_n la restriction de μ à E_n et ν_n la restriction de ν à E_n :

$$\forall A \in \mathcal{A} \ , \ \mu_n(A) = \mu(A \cap E_n), \ \nu_n(A) = \nu(A \cap E_n).$$

On peut appliquer la partie (1) à μ_n et ν_n, et on trouve $\mu_n = \nu_n$. Finalement, en utilisant les propriétés de limite croissante des mesures, pour tout $A \in \mathcal{A}$,

$$\mu(A) = \lim \uparrow \mu(A \cap E_n) = \lim \uparrow \nu(A \cap E_n) = \nu(A).$$

Conséquence. *Unicité de la mesure de Lebesgue.* Il existe au plus une mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tout intervalle ouvert non vide $]a, b[$, on ait $\lambda(]a, b[) = b - a$. En effet, si λ' est une seconde mesure ayant la même propriété, on peut appliquer à λ et λ' la partie (2) du corollaire précédent, en prenant pour $\mathcal{C}$ la classe des intervalles ouverts (dont on sait qu'elle engendre la tribu borélienne) et $E_n =]-n, n[$ pour tout n.

De la même façon, on déduit du corollaire précédent qu'une mesure *finie* μ sur $\mathbb{R}$ est caractérisée par les valeurs de $\mu(]-\infty, a])$ pour tout $a \in \mathbb{R}$.

Chapitre 2

Intégration par rapport à une mesure

Le premier objectif de ce chapitre est de construire l'intégrale de fonctions mesurables. La définition est facile pour les fonctions dites étagées, qui ne prennent qu'un nombre fini de valeurs. Ensuite l'intégrale d'une fonction mesurable positive est définie comme le supremum des intégrales des fonctions étagées qui la minorent. Pour les fonctions de signe quelconque, on raisonne par linéarité en se limitant aux fonctions dites intégrables, dont la valeur absolue est d'intégrale finie. Une fois construite l'intégrale, on établit les trois grands théorèmes de convergence de la théorie, à savoir le théorème de convergence monotone, le lemme de Fatou et le théorème de convergence dominée. Ces trois énoncés visent à donner des conditions assurant que l'intégrale de la limite d'une suite de fonctions est la limite des intégrales de ces fonctions. Le dernier paragraphe donne des applications importantes à la continuité ou la dérivabilité d'intégrales dépendant d'un paramètre.

2.1 Intégration de fonctions positives

On se donne un espace mesuré, c'est-à-dire un espace mesurable $(E, \mathcal{A})$ muni d'une mesure μ.

Fonctions étagées. Une fonction mesurable f à valeurs dans $\mathbb{R}$ est dite étagée si elle ne prend qu'un nombre fini de valeurs. Si $\alpha_1, \alpha_2, \ldots, \alpha_n$ sont les valeurs prises par f, qu'on peut supposer rangées par ordre croissant $\alpha_1 < \alpha_2 < \cdots < \alpha_n$, on a alors

$$f(x) = \sum_{i=1}^{n} \alpha_i \, 1_{A_i}(x)$$

où, pour chaque $i \in \{1, \ldots, n\}$, $A_i = f^{-1}(\{\alpha_i\}) \in \mathcal{A}$. L'écriture précédente sera appelée l'écriture canonique de f.

Définition 2.1.1 *Supposons f à valeurs dans $\mathbb{R}_+$. L'intégrale de f par rapport à μ est alors définie par*

$$\int f \, d\mu = \sum_{i=1}^{n} \alpha_i \, \mu(A_i)$$

avec la convention $0.\infty = 0$ dans le cas où $\alpha_i = 0$ et $\mu(A_i) = \infty$.

On a a priori $\int f d\mu \in [0, \infty]$.

Supposons qu'on ait une autre écriture de f sous la forme

$$f = \sum_{j=1}^{m} \beta_j \, 1_{B_j}$$

les ensembles mesurables B_j formant toujours une partition de E mais les nombres β_j n'étant plus nécessairement distincts. Alors il est facile de vérifier qu'on a aussi

$$\int f \, d\mu = \sum_{j=1}^{m} \beta_j \, \mu(B_j).$$

En effet, pour chaque $i \in \{1, \ldots, n\}$, A_i doit être la réunion disjointe des ensembles B_j pour les indices j tels que $\beta_j = \alpha_i$. Il suffit alors d'utiliser la propriété d'additivité de la mesure pour écrire

$$\mu(A_i) = \sum_{\{j:\beta_j=\alpha_i\}} \mu(B_j)$$

ce qui conduit au résultat annoncé.

Propriétés. Soient f et g deux fonctions étagées positives.

(1) Pour tous $a, b \geq 0$,

$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu.$$

(2) Si $f \leq g$,

$$\int f d\mu \leq \int g d\mu.$$

Preuve. (1) Soient

$$f = \sum_{i=1}^{n} \alpha_i \, 1_{A_i} \; , g = \sum_{k=1}^{m} \alpha'_k \, 1_{A'_k}$$

les écritures canoniques de f et g. En écrivant chaque A_i comme la réunion disjointe des ensembles $A_i \cap A'_k$, $k \in \{1, \ldots, m\}$, et de même pour chaque A'_k, on voit qu'on peut écrire

$$f = \sum_{j=1}^{p} \beta_j \, 1_{B_j} \; , \; g = \sum_{j=1}^{p} \gamma_j \, 1_{B_j}$$

avec les mêmes ensembles mesurables disjoints B_j (mais les nombres β_j, resp. γ_j, non nécessairement distincts). D'après la remarque suivant la définition, on a

$$\int f \, d\mu = \sum_{j=1}^{p} \beta_j \, \mu(B_j) \; , \; \int g \, d\mu = \sum_{j=1}^{p} \gamma_j \, \mu(B_j).$$

et de même $\int (af + bg) d\mu = \sum_{j=1}^{p} (a\beta_j + b\gamma_j) \, \mu(B_j)$, d'où le résultat voulu.

(2) On applique (1) en écrivant

$$\int g d\mu = \int f d\mu + \int (g - f) d\mu \geq \int f\, d\mu.$$

□

Notons $\mathcal{E}_+$ l'espace des fonctions étagées positives.

Définition 2.1.2 *Soit $f : E \longrightarrow [0, \infty]$ une fonction mesurable. On pose*

$$\int f\, d\mu = \sup_{h \in \mathcal{E}_+, h \leq f} \int h\, d\mu\,.$$

La propriété (2) ci-dessus montre que cette définition est cohérente avec la précédente quand f est étagée.

On notera indifféremment

$$\int f d\mu = \int f(x) d\mu(x) = \int f(x) \mu(dx)$$

et on trouve parfois la notation $\langle \mu, f \rangle$ ou même $\mu(f)$.

Propriétés.

(1) Si $f \leq g$, $\int f d\mu \leq \int g d\mu$ *(évident sur la définition)*

(2) Si $\mu(\{x \in E : f(x) > 0\}) = 0$, alors $\int f d\mu = 0$. *(en effet il suffit de le vérifier lorsque f est étagée, mais alors c'est évident sur la définition)*

Théorème 2.1.1 (Théorème de convergence monotone) *Soit (f_n) une suite croissante de fonctions mesurables positives (à valeurs dans $[0, \infty]$), et soit $f = \lim \uparrow f_n$. Alors*

$$\int f d\mu = \lim_{n \to \infty} \uparrow \int f_n d\mu.$$

Preuve. D'après la propriété (1) ci-dessus, on a

$$\int f d\mu \geq \lim_{n \to \infty} \uparrow \int f_n d\mu$$

et il suffit donc d'établir l'autre inégalité. Pour cela, choisissons une fonction étagée positive

$$h = \sum_{i=1}^{m} \alpha_i \, 1_{A_i}$$

avec $h \leq f$. Soit $a \in [0, 1[$, et

$$E_n = \{x \in E : ah(x) \leq f_n(x)\}.$$

Alors E_n est mesurable. De plus en utilisant le fait que f_n croît vers f, et la condition $a < 1$, on voit que E est la réunion croissante des ensembles E_n.

Ensuite on remarque qu'on a l'inégalité $f_n \geq a1_{E_n}h$, d'où

$$\int f_n d\mu \geq \int a1_{E_n} h \, d\mu = a \sum_{i=1}^{m} \alpha_i \, \mu(A_i \cap E_n).$$

Puisque $E_n \uparrow E$ on a $A_i \cap E_n \uparrow A_i$ et $\mu(A_i \cap E_n) \uparrow \mu(A_i)$ quand $n \to \infty$, d'après les propriétés élémentaires des mesures. En passant à la limite croissante il vient

$$\lim_{n\to\infty} \uparrow \int f_n d\mu \geq a \sum_{i=1}^{m} \alpha_i \, \mu(A_i) = a \int h d\mu.$$

En faisant tendre a vers 1, on trouve

$$\lim_{n\to\infty} \uparrow \int f_n d\mu \geq \int h d\mu.$$

Comme $\int f d\mu$ est définie par le supremum des quantités de droite lorsque h décrit l'ensemble des fonctions étagées positives majorées par f, on obtient bien l'inégalité recherchée. $\square$

Dans toute la suite "fonction mesurable positive" signifie fonction mesurable à valeurs dans $[0, \infty]$.

Proposition 2.1.2 (1) *Soit f une fonction mesurable positive. Il existe une suite croissante (f_n) de fonctions étagées positives telle que $f_n \uparrow f$.*

(2) *Si f et g sont mesurables positives et $a, b \in \mathbb{R}_+$,*

$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu.$$

(3) *Si (f_n) est une suite quelconque de fonctions mesurables positives,*

$$\int \sum_n f_n \, d\mu = \sum_n \int f_n \, d\mu.$$

Preuve. (1) Pour tout $n \geq 1$ et tout $i \in \{0, 1, \ldots, n2^n - 1\}$, posons

$$\begin{aligned} A_n &= \{x \in E : f(x) \geq n\} \\ B_{n,i} &= \{x \in E : i2^{-n} \leq f(x) < (i+1)2^{-n}\}. \end{aligned}$$

Soit ensuite f_n la fonction étagée

$$f_n = \sum_{i=0}^{n2^n-1} \frac{i}{2^n} \, 1_{B_{n,i}} + n \, 1_{A_n}.$$

On vérifie aisément que $f_n(x) \uparrow f(x)$ pour tout $x \in E$.

(2) On construit deux suites de fonctions étagées positives (f_n), (g_n) avec $f_n \uparrow f$, $g_n \uparrow g$. Alors on a aussi $af_n + bg_n \uparrow af + bg$, et en utilisant le théorème de convergence monotone et les propriétés de l'intégrale des fonctions étagées,

$$\int (af+bg)d\mu = \lim \uparrow \int (af_n+bg_n)d\mu = \lim \uparrow (a\int f_n d\mu + b\int g_n d\mu) = a\int f d\mu + b\int g d\mu.$$

(3) Cette assertion découle de (2) (cas d'une somme finie) et du théorème de convergence monotone. □

Remarque. Considérons le cas particulier où $E = \mathbb{N}$ et μ est la mesure de comptage. Alors il est facile de voir que

$$\int f\, d\mu = \sum_{k\in\mathbb{N}} f(k)$$

et (3) redonne la propriété bien connue énonçant que pour toute suite double $(a_{n,k})$ de réels positifs,

$$\sum_{k\in\mathbb{N}}\Big(\sum_{n\in\mathbb{N}} a_{n,k}\Big) = \sum_{n\in\mathbb{N}}\Big(\sum_{k\in\mathbb{N}} a_{n,k}\Big).$$

Corollaire 2.1.3 *Soit f mesurable positive, et pour tout $A \in \mathcal{A}$, soit*

$$\nu(A) = \int 1_A f\, d\mu \overset{\text{(not.)}}{=} \int_A f\, d\mu.$$

Alors ν est une mesure positive sur $(E, \mathcal{A})$, appelée mesure de densité f par rapport à μ, et notée $\nu = f \cdot \mu$.

Preuve. Il est immédiat que $\nu(\varnothing) = 0$. Par ailleurs, si (A_n) est une suite d'ensembles mesurables disjoints,

$$\nu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \int \sum_{n\in\mathbb{N}} 1_{A_n} f\, d\mu = \sum_{n\in\mathbb{N}} \int 1_{A_n} f d\mu = \sum_{n\in\mathbb{N}} \nu(A_n)$$

en utilisant la propriété (3) ci-dessus. □

Remarque. On a $\mu(A) = 0 \Rightarrow \nu(A) = \int 1_A f\, d\mu = 0$.

On dit qu'une propriété est vraie μ presque partout, ou μ p.p. ou même simplement p.p. s'il n'y a pas ambiguïté, si elle est vraie en dehors d'un ensemble de mesure nulle. Par exemple si f et g sont deux fonctions mesurables, $f = g$ p.p. signifie

$$\mu(\{x \in E : f(x) \neq g(x)\}) = 0.$$

Proposition 2.1.4 *Soit f une fonction mesurable positive.*

(1) Pour tout $a > 0$,

$$\mu(\{x \in E : f(x) \geq a\}) \leq \frac{1}{a} \int f d\mu.$$

(2) On a

$$\int f d\mu < \infty \Rightarrow f < \infty \quad p.p.$$

(3) On a

$$\int f d\mu = 0 \Leftrightarrow f = 0 \quad p.p.$$

(4) Si g est une autre fonction mesurable positive,

$$f = g \quad p.p. \Rightarrow \int f d\mu = \int g d\mu.$$

Preuve. (1) Posons $A_a = \{x \in E : f(x) \geq a\}$. Alors $f \geq a 1_{A_a}$ et donc

$$\int f d\mu \geq \int a 1_{A_a} \, d\mu = a\mu(A_a).$$

(2) Pour tout $n \geq 1$, soit $A_n = \{x \in E : f(x) \geq n\}$ et soit $A_\infty = \{x \in E : f(x) = \infty\}$. Alors, en utilisant (1),

$$\mu(A_\infty) = \mu\Big(\bigcap_{n\geq 1} A_n\Big) = \lim_{n\to\infty} \downarrow \mu(A_n) \leq \lim_{n\to\infty} \frac{1}{n} \int f d\mu = 0.$$

(3) L'implication $\Leftarrow$ a déjà été vue. Pour $\Rightarrow$, soit, pour tout $n \geq 1$, $B_n = \{x \in E : f(x) \geq n^{-1}\}$. Alors, d'après (1),

$$\mu(B_n) \leq n \int f d\mu = 0$$

et donc $\mu(B_n) = 0$ ce qui entraîne $\mu(\{x : f(x) > 0\}) = \mu\Big(\bigcup_{n\geq 1} B_n\Big) = 0$.

(4) Utilisons la notation $f \vee g = \sup(f, g)$ et $f \wedge g = \inf(f, g)$. Alors $f \vee g = f \wedge g$ p.p., d'où

$$\int (f \vee g) d\mu = \int (f \wedge g) d\mu + \int (f \vee g - f \wedge g) d\mu = \int (f \wedge g) d\mu,$$

puisque $f \vee g - f \wedge g = 0$ p.p. Puisque $f \wedge g \leq f \leq f \vee g$, et de même pour g, il en découle que

$$\int f d\mu = \int (f \vee g) d\mu = \int g d\mu.$$

Théorème 2.1.5 (Lemme de Fatou) *Soit (f_n) une suite quelconque de fonctions mesurables positives. Alors,*

$$\int (\liminf f_n) d\mu \leq \liminf \int f_n d\mu.$$

Preuve. On a

$$\liminf f_n = \lim_{k\to\infty} \uparrow \Big(\inf_{n\geq k} f_n\Big)$$

et donc d'après le théorème de convergence monotone,

$$\int (\liminf f_n) d\mu = \lim_{k\to\infty} \uparrow \int \Big(\inf_{n\geq k} f_n\Big) d\mu.$$

Par ailleurs, pour tout entier $p \geq k$,

$$\inf_{n\geq k} f_n \leq f_p$$

ce qui entraîne

$$\int \Big(\inf_{n\geq k} f_n\Big) d\mu \leq \inf_{p\geq k} \int f_p d\mu.$$

En passant à la limite croissante quand $k \uparrow \infty$, il vient

$$\lim_{k\to\infty} \uparrow \int \Big(\inf_{n\geq k} f_n\Big) d\mu \leq \lim_{k\to\infty} \uparrow \inf_{p\geq k} \int f_p d\mu = \liminf \int f_n d\mu,$$

ce qui termine la preuve. □

2.2 Fonctions intégrables

Définition 2.2.1 *Soit $f : E \longrightarrow \mathbb{R}$ une fonction mesurable. On dit que f est intégrable par rapport à μ (ou μ-intégrable) si*

$$\int |f| \, d\mu < \infty.$$

Dans ce cas on pose

$$\int f \, d\mu = \int f^+ d\mu - \int f^- d\mu$$

où $f^+ = \sup(f, 0)$, resp. $f^- = \sup(-f, 0)$ est la partie positive, resp. négative, de f. (Noter que f^+ et f^- sont mesurables et que $f = f^+ - f^-$ et $|f| = f^+ + f^-$.)

Remarque. On a $\int f^+ d\mu \leq \int |f| d\mu < \infty$ et de même $\int f^- d\mu < \infty$, ce qui montre que la définition de $\int f d\mu$ a bien un sens. Dans le cas où f est positive, cette définition coïncide bien sûr avec la précédente.

On note $\mathcal{L}^1(E, \mathcal{A}, \mu)$ l'espace des fonctions μ-intégrables. On utilisera parfois la notation $\mathcal{L}^1_+(E, \mathcal{A}, \mu)$ pour les fonctions μ-intégrables à valeurs positives.

Propriétés.

(a) $|\int f d\mu| \leq \int |f| d\mu$ pour $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$.

(b) $\mathcal{L}^1(E, \mathcal{A}, \mu)$ est un espace vectoriel et l'application $f \to \int f d\mu$ est une forme linéaire sur cet espace vectoriel.

(c) Si $f, g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ et $f \leq g$, alors $\int f d\mu \leq \int g d\mu$.

(d) Si $f,g \in \mathcal{L}^1(E,\mathcal{A},\mu)$ et $f = g$ μ p.p., alors $\int f d\mu = \int g d\mu$.

Preuve. (a) On écrit

$$|\int f d\mu| = |\int f^+ d\mu - \int f^- d\mu| \leq |\int f^+ d\mu| + |\int f^- d\mu| = \int |f| d\mu.$$

(b) Soit $f \in \mathcal{L}^1(E,\mathcal{A},\mu)$. Pour $a \in \mathbb{R}$,

$$\int |af| d\mu = |a| \int |f| d\mu < \infty.$$

Si $a \geq 0$,

$$\int (af) d\mu = \int (af)^+ d\mu - \int (af)^- d\mu = a \int f d\mu$$

et si $a < 0$,

$$\int (af) d\mu = \int (af)^+ d\mu - \int (af)^- d\mu = (-a) \int f^- d\mu + a \int f^+ d\mu = a \int f d\mu.$$

De plus, si $f,g \in \mathcal{L}^1(E,\mathcal{A},\mu)$, l'inégalité $|f+g| \leq |f| + |g|$ entraîne que $f + g \in \mathcal{L}^1$. En outre,

$$(f+g)^+ - (f+g)^- = f + g = f^+ - f^- + g^+ - g^-$$

entraîne

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+.$$

En utilisant l'additivité de l'intégrale pour les fonctions positives,

$$\int (f+g)^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int (f+g)^- d\mu + \int f^+ d\mu + \int g^+ d\mu,$$

d'où, puisque toutes les intégrales sont finies,

$$\int (f+g)^+ d\mu - \int (f+g)^- d\mu = \int f^+ d\mu - \int f^- d\mu + \int g^+ d\mu - \int g^- d\mu,$$

ce qui donne bien $\int (f+g) d\mu = \int f d\mu + \int g d\mu$.

(c) Il suffit d'écrire $\int g d\mu = \int f d\mu + \int (g - f) d\mu$.

(d) L'égalité $f = g$ p.p. entraîne $f^+ = g^+$ et $f^- = g^-$ p.p. Il suffit alors d'utiliser les résultats vus dans le cas des fonctions positives. $\square$

Remarque. On combine facilement (c) et (d) pour obtenir que, si $f,g \in \mathcal{L}^1(E,\mathcal{A},\mu)$ et $f \leq g$ p.p., alors $\int f d\mu \leq \int g d\mu$.

Extension au cas complexe. Soit $f : E \longrightarrow \mathbb{C}$ une fonction mesurable (cela équivaut à dire que $\mathrm{Re}(f)$ et $\mathrm{Im}(f)$ sont toutes deux mesurables). On dit que f est intégrable et on note $f \in \mathcal{L}^1_{\mathbb{C}}(E,\mathcal{A},\mu)$ si

$$\int |f| d\mu < \infty.$$

On pose alors

$$\int f\,d\mu = \int \mathrm{Re}(f)d\mu + i\int \mathrm{Im}(f)d\mu.$$

Les propriétés (a),(b) et (d) ci-dessus restent vraies si $\mathcal{L}^1(E,\mathcal{A},\mu)$ est remplacé par $\mathcal{L}^1_{\mathbb{C}}(E,\mathcal{A},\mu)$ (pour montrer (a), remarquer que

$$|\int f\,d\mu| = \sup_{a\in\mathbb{C},|a|=1} a\cdot \int f d\mu$$

où $a\cdot z$ désigne le produit scalaire dans $\mathbb{C}$ identifié à $\mathbb{R}^2$).

Théorème 2.2.1 (Théorème de convergence dominée) *Soit (f_n) une suite de fonctions dans $\mathcal{L}^1(E,\mathcal{A},\mu)$ (resp. dans $\mathcal{L}^1_{\mathbb{C}}(E,\mathcal{A},\mu)$). On suppose:*

(1) Il existe une fonction f mesurable à valeurs dans $\mathbb{R}$ (resp. dans $\mathbb{C}$) telle que

$$f_n(x) \longrightarrow f(x) \qquad \mu \text{ p.p.}$$

(2) Il existe une fonction $g : E \longrightarrow \mathbb{R}_+$ mesurable telle que $\int g d\mu < \infty$ et pour tout n,

$$|f_n| \le g \qquad \mu \text{ p.p.}$$

Alors $f \in \mathcal{L}^1(E,\mathcal{A},\mu)$ (resp. $f \in \mathcal{L}^1_{\mathbb{C}}(E,\mathcal{A},\mu)$), et on a

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$$

et

$$\lim_{n\to\infty} \int |f_n - f| d\mu = 0.$$

Preuve. On suppose d'abord que les hypothèses suivantes plus fortes sont vérifiées:

(1)' Pour tout $x \in E$,

$$f_n(x) \longrightarrow f(x)$$

(2)' Il existe une fonction $g : E \longrightarrow \mathbb{R}_+$ mesurable telle que $\int g d\mu < \infty$ et pour tout n et tout $x \in E$

$$|f_n(x)| \le g(x).$$

La propriété $f \in \mathcal{L}^1$ est alors claire puisque $|f| \le g$ et $\int g d\mu < \infty$. Ensuite, puisque $|f - f_n| \le 2g$ et $|f - f_n| \longrightarrow 0$, on peut appliquer le lemme de Fatou pour trouver

$$\liminf \int (2g - |f - f_n|)\,d\mu \ge \int \liminf (2g - |f - f_n|)\,d\mu = 2\int g d\mu.$$

Par linéarité de l'intégrale, il vient

$$2\int g d\mu - \limsup \int |f - f_n| d\mu \ge 2\int g d\mu,$$

d'où

$$\limsup \int |f - f_n| d\mu = 0,$$

et donc $\int |f - f_n| d\mu \longrightarrow 0$. Finalement il suffit d'écrire

$$\left| \int f d\mu - \int f_n d\mu \right| \leq \int |f - f_n| d\mu.$$

Dans le cas général où on suppose seulement (1) et (2), on pose

$$A = \{x \in E : f_n(x) \longrightarrow f(x) \text{ et pour tout } n,\ |f_n(x)| \leq g(x)\}.$$

Alors $\mu(A^c) = 0$, et on peut appliquer la première partie de la preuve aux fonctions

$$\tilde{f}_n(x) = 1_A(x) f_n(x)\ , \quad \tilde{f}(x) = 1_A(x) f(x).$$

On a $f = \tilde{f}$ p.p., $f_n = \tilde{f}_n$ p.p. et donc $\int f_n d\mu = \int \tilde{f}_n d\mu$, $\int f d\mu = \int \tilde{f} d\mu$ et $\int |f_n - f| d\mu = \int |\tilde{f}_n - \tilde{f}| d\mu$. Les résultats recherchés découlent du cas où (1)' et (2)' sont vérifiés. □

2.3 Intégrales dépendant d'un paramètre

On se donne un espace métrique (U, d) qui correspond à l'espace des paramètres. Soit une application $f : U \times E \longrightarrow \mathbb{R}$ (ou $\mathbb{C}$).

Théorème 2.3.1 *Soit $u_0 \in E$. Supposons*

(i) *pour tout $u \in U$, l'application $x \longrightarrow f(u, x)$ est mesurable;*

(ii) *$\mu(dx)$ p.p. l'application $u \longrightarrow f(u, x)$ est continue en u_0;*

(iii) *il existe une fonction $g \in \mathcal{L}^1_+(E, \mathcal{A}, \mu)$ telle que pour tout $u \in U$,*

$$|f(u, x)| \leq g(x) \qquad \mu(dx) \text{ p.p.}$$

Alors la fonction $F(u) = \int f(u, x) \mu(dx)$ est bien définie en tout point $u \in U$ et elle est continue en u_0.

Preuve. L'hypothèse (iii) entraîne que la fonction $x \longrightarrow f(u, x)$ est intégrable et donc $F(u)$ est bien définie. Ensuite, soit $(u_n)_{n \geq 1}$ une suite convergeant vers u_0. L'hypothèse (ii) assure que

$$f(u_n, x) \underset{n \to \infty}{\longrightarrow} f(u_0, x)\ , \quad \mu \text{ p.p.}$$

Grâce à l'hypothèse de domination (iii), on peut appliquer le théorème de convergence dominée, qui donne

$$\lim_{n \to \infty} \int f(u_n, x)\, \mu(dx) = \int f(u_0, x)\, \mu(dx).$$

□

Exemples. (a) Soit μ une mesure diffuse sur $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Si $\varphi\in\mathcal{L}^1(\mathbb{R},\mathcal{B}(\mathbb{R}),\mu)$, la fonction

$$F(u)=\int_{]-\infty,u]}\varphi(x)\,\mu(dx)=\int 1_{]-\infty,u]}(x)\varphi(x)\,\mu(dx)$$

est continue. Pour le voir, il suffit d'appliquer le théorème à $f(u,x)=1_{]-\infty,u]}(x)\varphi(x)$, en prenant $g=|\varphi|$ et en observant que pour $u_0\in\mathbb{R}$ fixé, la fonction $u\longrightarrow f(u,x)$ est continue en u_0 pour tout $x\in\mathbb{R}\backslash\{u_0\}$.

(b) Transformée de Fourier. Si $\varphi\in\mathcal{L}^1(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda)$, la fonction

$$\hat{\varphi}(u)=\int e^{iux}\,\varphi(x)\,\lambda(dx)$$

est continue sur $\mathbb{R}$.

(c) Convolution. Soit $\varphi\in\mathcal{L}^1(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda)$, et soit h une fonction continue bornée de $\mathbb{R}$ dans $\mathbb{R}$. Alors la fonction $h*\varphi$ définie sur $\mathbb{R}$ par

$$h*\varphi(u)=\int h(u-x)\,\varphi(x)\,\lambda(dx)$$

est continue (et bornée).

Nous passons maintenant à un théorème de dérivabilité sous le signe intégrale, et pour cela nous supposons que $U=I$ est un intervalle ouvert de $\mathbb{R}$. Soit à nouveau une application $f:U\times E\longrightarrow\mathbb{R}$ (ou $\mathbb{C}$).

Théorème 2.3.2 *Soit $u_0\in I$. Supposons que*

(i) *pour tout $u\in I$, l'application $x\longrightarrow f(u,x)$ est dans $\mathcal{L}^1(E,\mathcal{A},\mu)$;*

(ii) *$\mu(dx)$ p.p. l'application $u\longrightarrow f(u,x)$ est dérivable en u_0 de dérivée notée*

$$\frac{\partial f}{\partial u}(u_0,x)\ ;$$

(iii) *il existe une fonction $g\in\mathcal{L}^1_+(E,\mathcal{A},\mu)$ telle que pour tout $u\in I$,*

$$|f(u,x)-f(u_0,x)|\le g(x)|u-u_0|\qquad \mu(dx)\ p.p.$$

Alors la fonction $F(u)=\int f(u,x)\mu(dx)$ est dérivable en u_0, de dérivée

$$F'(u_0)=\int\frac{\partial f}{\partial u}(u_0,x)\,\mu(dx).$$

Remarque. A priori la dérivée $\frac{\partial f}{\partial u}(u_0,x)$ n'est définie (par (ii)) que pour x appartenant au complémentaire d'un ensemble de mesure nulle. On peut la prolonger à E tout entier de manière arbitraire (par exemple par la valeur 0), de façon à définir l'intégrale qui donne $F'(u_0)$.

Preuve. Soit $(u_n)_{n\geq 1}$ une suite dans $I\backslash\{u_0\}$ convergeant vers u_0, et soit

$$\varphi_n(x) = \frac{f(u_n,x) - f(u_0,x)}{u_n - u_0}.$$

Grâce à (ii), $\varphi_n(x)$ converge vers $\frac{\partial f}{\partial u}(u_0,x)$, $\mu(dx)$ p.p. De plus l'hypothèse (iii) permet d'appliquer le théorème de convergence dominée et d'obtenir

$$\lim_{n\to\infty} \frac{F(u_n) - F(u_0)}{u_n - u_0} = \lim_{n\to\infty} \int \varphi_n(x)\,\mu(dx) = \int \frac{\partial f}{\partial u}(u_0,x)\,\mu(dx).$$

□

Remarque. Dans de nombreuses applications, les hypothèses (ii) et (iii) sont remplacées par les hypothèses plus fortes

(ii)' $\mu(dx)$ p.p. l'application $u \longrightarrow f(u,x)$ est dérivable sur I;

(iii)' il existe une fonction $g \in \mathcal{L}^1_+(E,\mathcal{A},\mu)$ telle que $\mu(dx)$ p.p.,

$$\forall u \in I\ ,\quad \left|\frac{\partial f}{\partial u}(u,x)\right| \leq g(x).$$

(Noter que (iii)'⇒(iii) grâce au théorème des accroissements finis.) Sous ces hypothèses, la fonction F est dérivable sur I. L'exercice ci-dessous montre cependant que la forme plus précise de l'énoncé du théorème est parfois nécessaire.

Exemples. (a) Soit $\varphi \in \mathcal{L}^1(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda)$ telle que

$$\int |x\varphi(x)|\,\lambda(dx) < \infty.$$

Alors la transformée de Fourier $\hat{\varphi}(u)$ est dérivable sur $\mathbb{R}$, et

$$\hat{\varphi}'(u) = i\int x\,e^{iux}\,\varphi(x)\,\lambda(dx).$$

(b) Soit $\varphi \in \mathcal{L}^1(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda)$, et soit h une fonction de $\mathbb{R} \longrightarrow \mathbb{R}$ une fonction de classe C^1, bornée ainsi que sa dérivée. Alors la convolution $h * \varphi$ est dérivable sur $\mathbb{R}$, et

$$(h*\varphi)' = h' * \varphi.$$

On peut bien sûr itérer. Par exemple si h est de classe C^∞ à support compact, $h*\varphi$ est aussi de classe C^∞.

Exercice. Soit μ une mesure diffuse sur $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ et soit $\varphi \in \mathcal{L}^1(\mathbb{R},\mathcal{B}(\mathbb{R}),\mu)$ telle que

$$\int |x\varphi(x)|\,\mu(dx) < \infty.$$

Pour tout $u \in \mathbb{R}$, on pose

$$F(u) = \int_{\mathbb{R}} (u-x)^+\varphi(x)\,\mu(dx).$$

Montrer que F est dérivable sur $\mathbb{R}$, de dérivée

$$F'(u) = \int_{]-\infty,u]} \varphi(x)\,\mu(dx).$$

Chapitre 3

Construction de mesures

Le chapitre précédent partait de la donnée d'une mesure sur un espace mesurable. Nous montrons maintenant comment on construit des mesures intéressantes, et particulièrement la mesure de Lebesgue. Le premier paragraphe introduit la notion de mesure extérieure, vérifiant des propriétés des propriétés plus faibles que celles d'une mesure, et montre comment à partir d'une mesure extérieure on peut construire une (vraie) mesure sur une tribu convenable. Cette approche, qui est celle qu'avait utilisée Lebesgue, permet assez facilement de construire la mesure de Lebesgue sur $\mathbb{R}$ ou sur $\mathbb{R}^d$. Nous discutons aussi diverses propriétés de la mesure de Lebesgue, ainsi que ses liens avec l'intégrale de Riemann. Une autre application est l'intégrale de Stieltjes, qui correspond à l'intégrale par rapport à une mesure finie arbitraire sur la droite réelle.

3.1 Mesures extérieures

Définition 3.1.1 *Soit E un ensemble quelconque. Une application $\mu^* : \mathcal{P}(E) \longrightarrow [0, \infty]$ est appelée mesure extérieure si*

(i) $\mu^*(\varnothing) = 0$;

(ii) μ^* *est croissante :* $A \subset B \;\Rightarrow\; \mu^*(A) \leq \mu^*(B)$;

(iii) μ^* *est σ-sous-additive : pour toute suite A_k, $k \in \mathbb{N}$ d'éléments de $\mathcal{P}(E)$,*

$$\mu^*(\bigcup_{k\in\mathbb{N}} A_k) \leq \sum_{k\in\mathbb{N}} \mu^*(A_k).$$

Les propriétés d'une mesure extérieure sont moins contraignantes que celles d'une mesure. Remarquons cependant qu'une mesure extérieure est définie sur l'ensemble de toutes les parties de E et non pas seulement sur une tribu.

Nous verrons plus loin sur des exemples comment on construit des mesures extérieures. Notre objectif dans ce paragraphe est de montrer comment à partir d'une mesure extérieure μ^* on construit une mesure sur une tribu $\mathcal{M}(\mu^*)$ qui dépend de μ^*. Dans la suite de cette partie, on fixe une mesure extérieure μ^*.

Définition 3.1.2 *Une partie B de E est dite μ^*-mesurable si pour toute partie A de E,*

$$\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c).$$

On note $\mathcal{M}(\mu^*)$ l'ensemble des parties μ^*-mesurables.

Remarque. L'inégalité $\mu^*(A) \leq \mu^*(A \cap B) + \mu^*(A \cap B^c)$ est toujours vérifiée par σ-sous-additivité. Pour vérifier que B est μ^*-mesurable, c'est donc l'inégalité inverse qu'il importe de vérifier.

Théorème 3.1.1 (1) *$\mathcal{M}(\mu^*)$ est une tribu, qui contient toutes les parties B de E telles que $\mu^*(B) = 0$.*

(2) *La restriction de μ^* à $\mathcal{M}(\mu^*)$ est une mesure.*

Preuve. (1) Notons $\mathcal{M} = \mathcal{M}(\mu^*)$ pour simplifier. Si $\mu^*(B) = 0$, l'inégalité

$$\mu^*(A) \geq \mu^*(A \cap B^c) = \mu^*(A \cap B) + \mu^*(A \cap B^c)$$

montre aussitôt que $B \in \mathcal{M}$.

Ensuite on voit immédiatement que $\varnothing \in \mathcal{M}$ et que $\mathcal{M}$ est stable par passage au complémentaire. Pour terminer la preuve de la partie (1), il reste à montrer que $\mathcal{M}$ est stable par réunion dénombrable. On commence par établir que $\mathcal{M}$ est stable par réunion finie. Soient $B_1, B_2 \in \mathcal{M}$. Alors, pour toute $A \in \mathcal{P}(E)$, l'hypothèse $B_1 \in \mathcal{M}$ montre que

$$\mu^*(A \cap (B_1 \cup B_2)) = \mu^*(A \cap (B_1 \cup B_2) \cap B_1) + \mu^*(A \cap (B_1 \cup B_2) \cap B_1^c) = \mu^*(A \cap B_1) + \mu^*(A \cap B_2 \cap B_1^c).$$

Donc en utilisant successivement les propriétés $B_2 \in \mathcal{M}$ et $B_1 \in \mathcal{M}$,

$$\begin{aligned}
&\mu^*(A \cap (B_1 \cup B_2)) + \mu^*(A \cap (B_1 \cup B_2)^c) \\
&= \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c \cap B_2) + \mu^*(A \cap B_1^c \cap B_2^c) = \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c) = \mu^*(A),
\end{aligned}$$

ce qui montre bien que $B_1 \cup B_2 \in \mathcal{M}$. Etant stable par passage au complémentaire et par réunion finie, $\mathcal{M}$ est stable par intersection finie. En conséquence, si $B, B' \in \mathcal{M}$, $B \backslash B = B' \cap B^c \in \mathcal{M}$.

Compte-tenu de cette dernière remarque, il suffit pour compléter la preuve de montrer que si les ensembles $B_k \in \mathcal{M}$, $k \in \mathbb{N}$ sont deux à deux disjoints on a $\bigcup B_k \in \mathcal{M}$. Pour cela on montre par récurrence que pour tout entier $m \in \mathbb{N}$ et toute partie A de E,

$$\mu^*(A) = \sum_{k=0}^{m} \mu^*(A \cap B_k) + \mu^*(A \cap \Big(\bigcap_{k=0}^{m} B_k^c\Big)). \tag{3.1}$$

Pour $m = 0$, c'est la définition de $B_0 \in \mathcal{M}$. Pour passer de l'étape m à l'étape $m+1$, il suffit d'écrire

$$\begin{aligned}
\mu^*(A \cap \Big(\bigcap_{k=0}^{m} B_k^c\Big)) &= \mu^*(A \cap \Big(\bigcap_{k=0}^{m} B_k^c\Big) \cap B_{m+1}) + \mu^*(A \cap \Big(\bigcap_{k=0}^{m+1} B_k^c\Big)) \\
&= \mu^*(A \cap B_{m+1}) + \mu^*(A \cap \Big(\bigcap_{k=0}^{m+1} B_k^c\Big))
\end{aligned}$$

en utilisant le fait que les B_k sont disjoints. On déduit de (3.1) que

$$\mu^*(A) \geq \sum_{k=0}^{m} \mu^*(A \cap B_k) + \mu^*(A \cap \Big(\bigcap_{k=0}^{\infty} B_k^c\Big))$$

et en faisant tendre m vers ∞,

$$\begin{aligned}\mu^*(A) &\geq \sum_{k=0}^{\infty} \mu^*(A \cap B_k) + \mu^*(A \cap \Big(\bigcap_{k=0}^{\infty} B_k^c\Big)) \\ &\geq \mu^*(A \cap \Big(\bigcup_{k=0}^{\infty} B_k\Big)) + \mu^*(A \cap \Big(\bigcap_{k=0}^{\infty} B_k^c\Big)),\end{aligned}$$

par σ-sous-additivité. Cela suffit pour conclure que $\bigcup_{k=0}^{\infty} B_k \in \mathcal{M}$.

(2) Notons μ la restriction de μ^* à $\mathcal{M}$. On sait déjà que $\mu(\varnothing) = 0$. Soient B_k, $k \in \mathcal{M}$ des élements disjoints de $\mathcal{M}$. La preuve de (1) montre que pour toute partie A de E,

$$\mu^*(A) \geq \sum_{k=0}^{\infty} \mu^*(A \cap B_k) + \mu^*(A \cap \Big(\bigcap_{k=0}^{\infty} B_k^c\Big))$$

et donc en prenant $A = \bigcup_{k=0}^{\infty} B_k$,

$$\mu^*(\bigcup_{k=0}^{\infty} B_k) \geq \sum_{k=0}^{\infty} \mu^*(B_k).$$

Comme l'inégalité inverse est aussi vraie par σ-sous-additivité, cela termine la preuve. □

3.2 La mesure de Lebesgue

Pour toute partie A de $\mathbb{R}$, on définit

$$\lambda^*(A) = \inf\{\sum_{i\in\mathbb{N}} (b_i - a_i) : A \subset \bigcup_{i\in\mathbb{N}}]a_i, b_i[\}.$$

L'infimum porte sur tous les recouvrements dénombrables de A par des intervalles ouverts $]a_i, b_i[$, $a_i \leq b_i$ (évidemment il existe toujours de tels recouvrements).

Théorème 3.2.1 (i) *λ^* est une mesure extérieure sur $\mathbb{R}$.*

(ii) *La tribu $\mathcal{M}(\lambda^*)$ contient $\mathcal{B}(\mathbb{R})$.*

(iii) *Pour tous $a \leq b$, $\lambda^*([a, b]) = \lambda^*(]a, b[) = b - a$.*

La restriction de λ^* à $\mathcal{B}(\mathbb{R})$ (ou à $\mathcal{M}(\lambda^*)$) est la **mesure de Lebesgue** sur $\mathbb{R}$, et sera notée simplement λ. En conséquence des résultats de la fin du Chapitre 1, c'est l'unique mesure sur $\mathcal{B}(\mathbb{R})$ qui vérifie la propriété $\lambda(]a, b[) = b - a$ pour tout intervalle $]a, b[$.

Preuve. (i) Il est immédiat que $\lambda^*(\varnothing) = 0$ et que λ^* est croissante. Il reste à établir la sous-additivité. Pour cela, on se donne une suite $(A_n)_{n\in\mathbb{N}}$ de parties de $\mathbb{N}$. On peut supposer $\lambda^*(A_n) < \infty$ pour tout n (sinon il n'y a rien à montrer). Soit $\varepsilon > 0$. Pour tout $n \in \mathbb{N}$, on peut trouver une suite d'intervalles $]a_i^{(n)}, b_i^{(n)}[$, $i \in \mathbb{N}$ tels que

$$A_n \subset \bigcup_{i\in\mathbb{N}}]a_i^{(n)}, b_i^{(n)}[$$

et

$$\sum_{i\in\mathbb{N}} (b_i^{(n)} - a_i^{(n)}) \leq \lambda^*(A_n) + \frac{\varepsilon}{2^i}.$$

Il suffit alors de remarquer que les intervalles $]a_i^{(n)}, b_i^{(n)}[, n \in \mathbb{N}, i \in \mathbb{N}$ forment un recouvrement dénombrable de la réunion des A_n, et donc

$$\lambda^*(\bigcup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} \sum_{i\in\mathbb{N}} (b_i^{(n)} - a_i^{(n)}) \leq \sum_{n\in\mathbb{N}} \lambda^*(A_n) + 2\varepsilon,$$

d'où le résultat puisque ε est arbitraire.

(ii) Puisque $\mathcal{M}(\lambda^*)$ est une tribu, il suffit de montrer qu'elle contient une famille qui engendre la tribu borélienne, par exemple la famille des intervalles $] - \infty, \alpha]$, $\alpha \in \mathbb{R}$. On se donne donc $\alpha \in \mathbb{R}$ et on pose $B =] - \infty, \alpha]$. Le problème est de vérifier que pour toute partie A de $\mathbb{R}$,

$$\lambda^*(A) \geq \lambda^*(A \cap B) + \lambda^*(A \cap B^c).$$

Soit $(]a_i, b_i[)_{i\in\mathbb{N}}$ un recouvrement de A, et $\varepsilon > 0$. Les intervalles $]a_i \wedge \alpha, (b_i \wedge \alpha) + \varepsilon 2^{-i}[$ recouvrent $A \cap B$, et les intervalles $]a_i \vee \alpha, b_i \vee \alpha[$ recouvrent $A \cap B^c$. Donc

$$\lambda^*(A \cap B) \leq \sum_{i\in\mathbb{N}} ((b_i \wedge \alpha) - (a_i \wedge \alpha)) + 2\varepsilon,$$

$$\lambda^*(A \cap B^c) \leq \sum_{i\in\mathbb{N}} ((b_i \vee \alpha) - (a_i \vee \alpha)).$$

En faisant la somme on trouve

$$\lambda^*(A \cap B) + \lambda^*(A \cap B^c) \leq \sum_{i\in\mathbb{N}} (b_i - a_i) + 2\varepsilon.$$

Puisque ε était arbitraire, on a

$$\lambda^*(A \cap B) + \lambda^*(A \cap B^c) \leq \sum_{i\in\mathbb{N}} (b_i - a_i),$$

et comme $\lambda^*(A)$ est par définition l'infimum des sommes de droite sur tous les recouvrements de A, l'inégalité recherchée en découle.

(iii) Il est immédiat par définition que

$$\lambda^*([a,b]) \leq b-a.$$

Pour l'inégalité inverse, supposons que

$$[a,b] \subset \bigcup_{i\in\mathbb{N}}]a_i, b_i[.$$

Par compacité, on peut trouver un entier N assez grand tel que

$$[a,b] \subset \bigcup_{i=0}^{N}]a_i, b_i[.$$

Un raisonnement élémentaire montre alors que

$$b-a \leq \sum_{i=0}^{N}(b_i-a_i) \leq \sum_{i=0}^{\infty}(b_i-a_i).$$

Cela donne l'autre inégalité $b-a \leq \lambda^*([a,b])$. Il est facile de voir enfin que $\lambda^*(]a,b[) = \lambda^*([a,b])$ (par exemple en observant que $\lambda^*(\{a\}) = \lambda^*(\{b\}) = 0$). □

Extension en dimension d.

On appelle pavé ouvert (resp. fermé) un sous-ensemble P de $\mathbb{R}^d$ de la forme

$$P = \prod_{j=1}^{d}]a_j, b_j[\quad , \quad (\text{resp. } P = \prod_{j=1}^{d} [a_j, b_j]).$$

Le volume de P est par définition

$$\mathrm{vol}\,(P) = \prod_{j=1}^{d} (b_j - a_j).$$

On définit alors pour toute partie A de $\mathbb{R}^d$

$$\lambda^*(A) = \inf\{\sum_{i\in\mathbb{N}} \mathrm{vol}\,(P_i) : A \subset \bigcup_{i\in\mathbb{N}} P_i\}.$$

où l'infimum porte sur tous les recouvrements dénombrables de A par des pavés ouverts.

On a alors l'analogue suivant du théorème précédent.

Théorème 3.2.2 *(i) λ^* est une mesure extérieure sur $\mathbb{R}^d$.*

(ii) La tribu $\mathcal{M}(\lambda^*)$ contient $\mathcal{B}(\mathbb{R}^d)$.

(iii) Pour tous pavé (ouvert ou fermé) P, $\lambda^*(P) = \mathrm{vol}\,(P)$.

La restriction de λ^* à $\mathcal{B}(\mathbb{R}^d)$ (ou à $\mathcal{M}(\lambda^*)$) est la **mesure de Lebesgue** sur $\mathbb{R}^d$, et sera notée simplement λ.

Preuve. La preuve de (i) est exactement la même que dans le cas $d = 1$. Pour (ii), il suffit de montrer que si A est un ensemble de la forme

$$A = \mathbb{R} \times \cdots \times \mathbb{R} \times]-\infty, a] \times \mathbb{R} \times \cdots \times \mathbb{R},$$

on a $A \in \mathcal{M}(\lambda^*)$ (il est facile de voir que les ensembles de cette forme engendrent la tribu $\mathcal{B}(\mathbb{R}^d)$). La démonstration est alors tout à fait semblable à celle du cas $d = 1$. Enfin pour (iii), on se ramène à montrer que si P est un pavé fermé et si

$$P \subset \bigcup_{i=1}^{n} P_i$$

où les P_i sont des pavés ouverts, on a

$$\mathrm{vol}\,(P) \leq \sum_{i=1}^{n} \mathrm{vol}\,(P_i).$$

Cette assertion est laissée en exercice. □

Remarque. On verra plus tard (dans le Chapitre 5) une autre façon de construire la mesure de Lebesgue en dimension d à partir du cas de la dimension un.

On peut se demander si la tribu $\mathcal{M}(\lambda^*)$ est beaucoup plus grande que la tribu $\mathcal{B}(\mathbb{R})$. Nous allons voir qu'en un certain sens ces deux tribus ne sont pas très différentes. Nous énonçons d'abord une proposition préliminaire.

Proposition 3.2.3 *Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. La classe des parties négligeables est par définition*

$$\mathcal{N} = \{A \in \mathcal{P}(E) : \exists B \in \mathcal{A}, A \subset B \text{ et } \mu(B) = 0\}.$$

La tribu complétée de $\mathcal{A}$ (par rapport à μ) est $\bar{\mathcal{A}} = \sigma(\mathcal{A} \cup \mathcal{N})$. Il existe alors une unique mesure sur $(E, \bar{\mathcal{A}})$ qui prolonge μ.

Preuve. On remarque d'abord que la tribu $\bar{\mathcal{A}}$ peut être obtenue de la manière suivante : si

$$\mathcal{B} = \{A \in \mathcal{P}(E) : \exists B, B' \in \mathcal{A}, B \subset A \subset B' \text{ et } \mu(B' \backslash B) = 0\}$$

on a $\bar{\mathcal{A}} = \mathcal{B}$. En effet on vérifie facilement que $\mathcal{B}$ est une tribu. Il est clair que $\mathcal{A} \subset \mathcal{B}$ et $\mathcal{N} \subset \mathcal{B}$, ce qui entraîne que $\bar{\mathcal{A}} \subset \mathcal{B}$. Enfin, si $A \in \mathcal{B}$, on choisit B et B' comme dans la définition et on remarque que $A = B \cup (A \backslash B)$, avec $B \in \mathcal{A}$ et $A \backslash B \in \mathcal{N}$. L'inclusion $\mathcal{B} \subset \bar{\mathcal{A}}$ en découle.

Une fois acquise l'égalité $\bar{\mathcal{A}} = \mathcal{B}$, on construit le prolongement de μ à $\bar{\mathcal{A}}$ de la manière suivante. Si $A \in \bar{\mathcal{A}} = \mathcal{B}$, et si B et B' sont comme dans la définition de $\mathcal{B}$ ci-dessus, on pose $\mu(A) = \mu(B) = \mu(B')$. Cela ne dépend pas du choix de B et B' : si $\tilde{B}$, $\tilde{B}'$ est un autre choix, on a à la fois $\mu(\tilde{B}) \leq \mu(B')$ et $\mu(\tilde{B}') \geq \mu(B)$ ce qui force les égalités

$\mu(B) = \mu(B') = \mu(\tilde{B}) = \mu(\tilde{B}')$. Enfin, il est facile de vérifier que le prolongement de μ à $\bar{\mathcal{A}}$ est une mesure : si A_n, $n \in \mathbb{N}$ sont des éléments disjoints de $\bar{\mathcal{A}}$, on peut pour chaque n choisir $B_n \in \mathcal{A}$, $B_n \subset A_n$ de manière que $A_n \backslash B_n$ soit négligeable, et on a

$$\sum_n \mu(A_n) = \sum_n \mu(B_n) = \mu(\bigcup_n B_n) = \mu(\bigcup_n A_n),$$

la dernière égalité parce que $\bigcup_n A_n \backslash \bigcup_n B_n \subset \bigcup_n (A_n \backslash B_n)$ est négligeable. □

Proposition 3.2.4 *La tribu $\mathcal{M}(\lambda^*)$ coïncide avec la complétée $\bar{\mathcal{B}}(\mathbb{R}^d)$ de $\mathcal{B}(\mathbb{R}^d)$ par rapport à la mesure de Lebesgue λ.*

Preuve. L'inclusion $\bar{\mathcal{B}}(\mathbb{R}^d) \subset \mathcal{M}(\lambda^*)$ est immédiate : si $A \in \mathcal{P}(\mathbb{R}^d)$ est tel que $A \subset B$, où $B \in \mathcal{B}(\mathbb{R}^d)$ et $\lambda(B) = 0$, alors $\lambda^*(A) \leq \lambda^*(B) = \lambda(B) = 0$, et d'après le théorème du paragraphe 1, on sait que cela entraîne $A \in \mathcal{M}(\lambda^*)$.

Inversement, soit $A \in \mathcal{M}(\lambda^*)$. On veut montrer que $A \in \bar{\mathcal{B}}(\mathbb{R}^d)$. Sans perte de généralité, on peut supposer $A \subset]-K, K[^d$ (sinon on écrit A comme la réunion croissante des ensembles $A \cap]-n, n[^d$). On a alors $\lambda^*(A) < \infty$, et donc pour chaque $n \geq 1$ on peut trouver une famille dénombrable $(P_i^n, i \in \mathbb{N})$ de pavés ouverts contenus dans $]-K, K[^d$ tels que

$$A \subset \bigcup_i P_i^n \ , \qquad \sum_i \mathrm{vol}\,(P_i^n) \leq \lambda^*(A) + \frac{1}{n}.$$

Posons

$$B_n = \bigcup_i P_i^n, \qquad B = \bigcap_n B_n.$$

Alors $B \in \mathcal{B}(\mathbb{R}^d)$, $A \subset B$, et d'autre part pour chaque n,

$$\lambda^*(B) \leq \sum_i \mathrm{vol}\,(P_i^n) \leq \lambda^*(A) + \frac{1}{n}$$

ce qui implique $\lambda^*(B) = \lambda^*(A)$. En remplaçant A par $]-K, K[^d \backslash A$, on construit de même $\tilde{B} \in \mathcal{B}(\mathbb{R}^d)$, $\tilde{B} \subset]-K, K[^d$ telle que $]-K, K[^d \backslash A \subset \tilde{B}$ et $\lambda^*(]-K, K[^d \backslash A) = \lambda^*(\tilde{B})$. Si $B' =]-K, K[^d \backslash \tilde{B}$, on doit alors avoir $B' \subset A$ et $\lambda^*(B') = \lambda^*(A)$. Finalement on a bien trouvé deux boréliens B et B' avec $B' \subset A \subset B$ et $\lambda(B \backslash B') = 0$. □

Théorème 3.2.5 *La mesure de Lebesgue sur $\mathbb{R}^d$ est invariante par translation, au sens où pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ et tout $x \in \mathbb{R}^d$, on a $\lambda(x + A) = \lambda(A)$.*

Inversement, si μ est une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ finie sur les parties bornées et invariante par translation, il existe une constante $c \geq 0$ telle que $\mu = c\lambda$.

Preuve. Notons σ_x la translation $\sigma_x(y) = y - x$ pour tout $y \in \mathbb{R}^d$. La mesure-image $\sigma_x(\lambda)$ est définie par

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \sigma_x(\lambda)(A) = \lambda(\sigma_x^{-1}(A)) = \lambda(x + A).$$

L'égalité $\sigma_x(\lambda)(A) = \lambda(A)$ est vraie pour tout pavé A (puisque A et $x+A$ sont deux pavés de même volume). A l'aide du lemme de classe monotone du Chapitre 1, il en découle aussitôt que $\sigma_x(\lambda) = \lambda$, ce qui est la première assertion du théorème.

Inversement, soit μ une mesure sur $\mathcal{B}(\mathbb{R}^d)$ invariante par translation. Soit

$$c = \mu([0,1[^d).$$

Comme $[0,1[^d$ est la réunion disjointe de n^d pavés qui sont des translatés de $[0,\frac{1}{n}[^d$, il en résulte que pour tout entier $n \geq 1$,

$$\mu([0,\frac{1}{n}[^d) = \frac{c}{n^d}.$$

Soient ensuite $a_1,\ldots,a_d \geq 0$. En écrivant

$$\prod_{j=1}^{d}[0,\frac{[na_j]}{n}[\subset \prod_{j=1}^{d}[0,a_j[\subset \prod_{j=1}^{d}[0,\frac{[na_j]+1}{n}[$$

(où $[x]$ désigne la partie entière de x), on trouve

$$(\prod_{j=1}^{d}[na_j])\frac{c}{n^d} = \mu(\prod_{j=1}^{d}[0,\frac{[na_j]}{n}[) \leq \mu(\prod_{j=1}^{d}[0,a_j[) \leq \mu(\prod_{j=1}^{d}[0,\frac{[na_j]+1}{n}[) = (\prod_{j=1}^{d}[na_j]+1)\frac{c}{n^d}.$$

En faisant tendre n vers ∞, il vient

$$\mu(\prod_{j=1}^{d}[0,a_j[) = c\prod_{j=1}^{n}a_j = c\lambda(\prod_{j=1}^{d}[0,a_j[)$$

et en utilisant l'invariance par translation de μ on trouve que les mesures μ et $c\lambda$ coïncident sur tous les pavés de la forme

$$\prod_{j=1}^{d}[a_j,b_j[.$$

Comme dans la première partie de la preuve, cela suffit pour conclure que $\mu = c\lambda$. □

Proposition 3.2.6 *La mesure de Lebesgue sur $\mathbb{R}^d$ est régulière au sens où pour tout $A \in \bar{B}(\mathbb{R}^d)$, on a*

$$\begin{aligned}\lambda(A) &= \inf\{\lambda(U) : U \textit{ ouvert }, A\subset U\}\\ &= \sup\{\lambda(F) : F \textit{ compact }, F\subset A\}.\end{aligned}$$

Preuve. La quantité $\inf\{\lambda(U) : U \text{ ouvert }, A\subset U\}$ est toujours plus grande que $\lambda(A)$. Pour l'autre inégalité, on peut supposer $\lambda(A) < \infty$. Ensuite, par définition de $\lambda(A) = \lambda^*(A)$, on peut pour chaque $\varepsilon > 0$ trouver un recouvrement de A par des pavés ouverts P_i tels que $\sum \lambda(P_i) \leq \lambda(A) + \varepsilon$. Mais alors l'ouvert U défini comme la réunion des P_i contient A et on a $\lambda(U) \leq \sum \lambda(P_i) \leq \lambda(A) + \varepsilon$, ce qui conduit à l'inégalité voulue.

Pour la deuxième égalité de la proposition, on peut supposer A contenu dans un compact C (sinon on écrit $\lambda(A) = \lim \uparrow \lambda(A \cap [-n,n]^d)$). Pour chaque $\varepsilon > 0$ on peut grâce à la première partie de la preuve trouver un ouvert U contenant $C\backslash A$, tel que $\lambda(U) < \lambda(C\backslash A)+\varepsilon$. Mais alors $F = C\backslash U$ est un compact contenu dans A, et

$$\lambda(F) \geq \lambda(C) - \lambda(U) \geq \lambda(C) - (\lambda(C\backslash A) + \varepsilon) = \lambda(A) - \varepsilon,$$

ce qui donne la deuxième égalité. □

La proposition précédente peut être étendue à un cadre beaucoup plus général. Nous nous limitons au cas des mesures finies.

Proposition 3.2.7 *Soit (E,d) un espace métrique, et soit μ une mesure finie sur $(E,\mathcal{B}(E))$. Alors, pour tout $A \in \mathcal{B}(E)$,*

$$\begin{aligned} \mu(A) &= \inf\{\mu(U) : U \text{ ouvert }, A \subset U\} \\ &= \sup\{\mu(F) : F \text{ fermé }, F \subset A\}. \end{aligned}$$

Preuve. Notons $\mathcal{O}$ la classe des ouverts de E, et soit $\mathcal{C}$ la classe des ensembles $A \in \mathcal{B}(E)$ qui vérifient la propriété de la proposition. Puisque la tribu borélienne est par définition engendrée par $\mathcal{O}$, il suffit de montrer que $\mathcal{O} \subset \mathcal{C}$ et que $\mathcal{C}$ est une tribu.

Si $A \in \mathcal{O}$, la première égalité est triviale. Pour la seconde, on remarque que pour tout $n \geq 1$, l'ensemble

$$F_n = \{x \in E : d(x, A^c) \geq \frac{1}{n}\}$$

est fermé. Par ailleurs $A = \lim \uparrow F_n$, ce qui entraîne

$$\mu(A) = \lim \uparrow \mu(F_n),$$

ce qui donne bien la seconde égalité et prouve que $\mathcal{O} \subset \mathcal{C}$.

Il reste à montrer que $\mathcal{C}$ est une tribu. On a $\varnothing \in \mathcal{C}$ et à cause de la symétrie entre ouverts et fermés, on voit immédiatement que $\mathcal{C}$ est stable par passage au complémentaire. Soit ensuite $(A_n)_{n\in\mathbb{N}}$ une suite dans $\mathcal{C}$ et soit $\varepsilon > 0$. Pour chaque n, on peut trouver un ouvert U_n contenant A_n tel que $\mu(U_n) \leq \mu(A_n) + \varepsilon 2^{-n}$, d'où

$$\mu\Big(\bigcup_{n\in\mathbb{N}} U_n \backslash \bigcup_{n\in\mathbb{N}} A_n\Big) \leq \sum_{n\in\mathbb{N}} \mu(U_n - A_n) \leq 2\varepsilon.$$

Puisque $\bigcup U_n$ est ouvert cela donne la première des deux égalités recherchées pour $\bigcup A_n$.

Ensuite, soit N un entier assez grand pour que

$$\mu(\bigcup_{n=0}^{N} A_n) \geq \mu(\bigcup_{n\in\mathbb{N}} A_n) - \varepsilon.$$

Pour chaque $n \in \{0,1,\ldots,N\}$ on peut trouver un fermé $F_n \subset A_n$ tel que $\mu(A_n\backslash F_n) \leq \varepsilon 2^{-n}$. Alors

$$F = \bigcup_{n=0}^{N} F_n$$

est fermé et

$$\mu((\bigcup_{n=0}^{N} A_n)\backslash F) \leq \sum_{n=0}^{N} \mu(A_n - F_n) < 2\varepsilon$$

d'où

$$\mu((\bigcup_{n=0}^{\infty} A_n)\backslash F) \leq 3\varepsilon.$$

On conclut que $\bigcup A_n \in \mathcal{C}$, ce qui termine la preuve.

3.3 Liens avec l'intégrale de Riemann

Fixons un intervalle $[a, b]$ non trivial de $\mathbb{R}$. Une fonction $h : [a, b] \longrightarrow \mathbb{R}$ est dite en escalier, et on note $h \in \mathrm{Esc}$, s'il existe une subdivision $a = x_0 < x_1 < \cdots < x_N = b$ et des réels $y_1, \ldots, y_N$ tels que

$$\forall i \in \{1, \ldots, N\},\ \forall x \in]x_{i-1}, x_i[,\ f(x) = y_i.$$

On pose alors

$$I(h) = \sum_{i=1}^{N} y_i(x_i - x_{i-1}).$$

Il est immédiat que $I(h) = \int_{[a,b]} h\, d\lambda$.

Une fonction bornée $f : [a, b] \longrightarrow \mathbb{R}$ est dite Riemann-intégrable si

$$\sup_{h \in \mathrm{Esc},\, h \leq f} I(h) = \inf_{h \in \mathrm{Esc},\, h \geq f} I(h)$$

et cette valeur commune est notée $I(f)$.

Proposition 3.3.1 *Soit f une fonction Riemann-intégrable sur $[a, b]$. Alors f est mesurable pour la tribu complétée $\bar{\mathcal{B}}([a, b])$, et*

$$I(f) = \int_{[a,b]} f\, d\lambda.$$

Preuve. On peut trouver une suite (h_n) de fonctions en escalier sur $[a, b]$ telles que $h_n \geq f$ et $I(h_n) \downarrow I(f)$. Quitte à remplacer h_n par $h_1 \wedge h_2 \wedge \cdots \wedge h_n$, on peut supposer la suite (h_n) décroissante, ce qui permet de poser

$$h_\infty = \lim \downarrow h_n \geq f.$$

De même, on peut trouver une suite croissante $(\tilde{h}_n)$ de fonctions en escalier avec $\tilde{h}_n \leq f$ et $I(\tilde{h}_n) \uparrow I(f)$, et poser

$$\tilde{h}_\infty = \lim \uparrow \tilde{h}_n \leq f.$$

Les fonctions h_∞ et $\tilde{h}_\infty$ sont boréliennes bornées. Par convergence dominée,

$$\int_{[a,b]} h_\infty d\lambda = \lim \downarrow \int_{[a,b]} h_n d\lambda = \lim \downarrow I(h_n) = I(f),$$
$$\int_{[a,b]} \tilde{h}_\infty d\lambda = \lim \uparrow \int_{[a,b]} \tilde{h}_n d\lambda = \lim \uparrow I(\tilde{h}_n) = I(f).$$

Donc,

$$\int_{[a,b]} (h_\infty - \tilde{h}_\infty) d\lambda = 0.$$

Puisque $h_\infty \geq \tilde{h}_\infty$, cela entraîne $h_\infty = \tilde{h}_\infty$, λ p.p. Comme $h_\infty \geq f \geq \tilde{h}_\infty$, f coïncide p.p. avec une fonction borélienne, et il est facile d'en déduire que f est $\bar{\mathcal{B}}([a,b])$-mesurable. Enfin puisque $f = h_\infty$ p.p. on a $\int_{[a,b]} f\, d\lambda = \int_{[a,b]} h_\infty\, d\lambda = I(f)$. □

3.4 Un exemple d'ensemble non mesurable

Considérons l'espace $\mathbb{R}/\mathbb{Q}$ des classes d'équivalence des réels modulo les rationnels. Pour chaque $a \in \mathbb{R}/\mathbb{Q}$, soit x_a un représentant de a dans l'intervalle $[0,1]$. On pose

$$F = \{x_a\,;\, a \in \mathbb{R}/\mathbb{Q}\} \subset [0,1].$$

Alors F n'est pas borélien, ni même mesurable par rapport à la tribu complétée $\bar{\mathcal{B}}(\mathbb{R})$.

Pour le vérifier, supposons F mesurable et montrons que cela conduit à une contradiction. D'abord, on a par construction

$$\mathbb{R} \subset \bigcup_{q \in \mathbb{Q}} (q + F)$$

et donc $\lambda(F) > 0$, car sinon $\mathbb{R}$ serait contenu dans une réunion dénombrable d'ensembles de mesure nulle.

Par ailleurs, les ensembles $q + F$, $q \in \mathbb{Q}$ sont disjoints (si $q + x_a = q' + x_{a'}$ on a $x_a - x_{a'} = q' - q \in \mathbb{Q}$ et donc $a = a'$ puis $q = q'$). De l'inclusion

$$\bigcup_{q \in \mathbb{Q} \cap [0,1]} (q + F) \subset [0,2]$$

on déduit donc

$$\sum_{q \in \mathbb{Q} \cap [0,1]} \lambda(q + F) \leq 2$$

d'où $\lambda(F) = 0$ ce qui est la contradiction recherchée.

3.5 Intégrale de Stieltjes

Le théorème suivant donne une description de toutes les mesures finies sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Le résultat peut être facilement étendu aux mesures de Radon.

Théorème 3.5.1 (i) *Soit μ une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Pour tout $x \in \mathbb{R}$, soit*

$$F_\mu(x) = \mu(]-\infty, x]).$$

La fonction F_μ est croissante, bornée, continue à droite et $F_\mu(-\infty) = 0$.
(ii) *Inversement, soit $F : \mathbb{R} \longrightarrow \mathbb{R}_+$ une fonction est croissante, bornée, continue à droite et telle que $F(-\infty) = 0$. Il existe alors une unique mesure finie μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $F = F_\mu$.*

Remarque. Lorsque $F = F_\mu$, on note souvent

$$\int f(x)\,\mu(dx) = \int f(x)\,dF(x).$$

C'est l'intégrale de Stieltjes de f par rapport à F. On a en particulier

$$\int_{]a,b]} dF(x) = F(b) - F(a),$$

et

$$\int_{[a,b]} dF(x) = \lim_{n\to\infty} \int_{]a-n^{-1},b]} dF(x) = F(b) - F(a-),$$

où $F(a-)$ désigne la limite à gauche de F en a.

Preuve. (i) La vérification des propriétés de F_μ est facile. Par exemple si $x_n \downarrow x$, les intervalles $]-\infty, x_n]$ décroissent vers $]-\infty, x]$, et donc

$$F_\mu(x_n) = \mu(]-\infty, x_n]) \downarrow \mu(]-\infty, x]) = F(x).$$

De même, si $x_n \downarrow -\infty$, les intervalles $]-\infty, x_n]$ décroissent vers $\varnothing$ et donc $F_\mu(x_n) \downarrow 0$.
(ii) L'unicité de μ est une conséquence du lemme de classe monotone (cf Chapitre 1) : la classe $\mathcal{C} = \{]-\infty, x]; x \in \mathbb{R}\}$ est stable par intersection finie et engendre la tribu $\mathcal{B}(\mathbb{R})$.

Pour montrer l'existence, on pose pour tout $A \subset \mathbb{R}$:

$$\mu^*(A) = \inf\{\sum_{i\in\mathbb{N}} (F(b_i) - F(a_i)) : A \subset \bigcup_{i\in\mathbb{N}}]a_i, b_i]\}.$$

(Noter qu'on recouvre A par des intervalles ouverts à droite et fermés à gauche, et non plus des intervalles ouverts comme pour la mesure de Lebesgue.) Les mêmes arguments que dans le cas de la mesure de Lebesgue montrent que μ^* est une mesure extérieure. On vérifie par la même méthode que dans le cas de la mesure de Lebesgue que les intervalles $]-\infty, \alpha]$ sont dans $\mathcal{M}(\lambda^*)$ (en fait c'est même plus facile ici). Il en découle que la tribu $\mathcal{M}(\mu^*)$ contient la tribu borélienne, et que la restriction, notée μ, de μ^* à $\mathcal{M}(\mu^*)$ est une mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Pour terminer, il reste à montrer que $\mu(]-\infty, x]) = F(x)$ pour tout $x \in \mathbb{R}$. Il suffit pour cela d'établir que $\mu(]a, b]) = F(b) - F(a)$ pour tous $a < b$ (ensuite faire tendre a vers $-\infty$). L'inégalité

$$\mu(]a, b]) \leq F(b) - F(a)$$

est immédiate par construction de μ^*.

Dans l'autre sens, soit $(]x_i, y_i])_{i\in\mathbb{N}}$ un recouvrement dénombrable de $]a, b]$. Soit $\varepsilon \in]0, b-a[$. Pour chaque $i \in \mathbb{N}$, on peut trouver $y'_i > y_i$ tel que $F(y'_i) \leq F(y_i) + \varepsilon 2^{-i}$. Ensuite, on remarque qu'on peut recouvrir l'intervalle compact $[a+\varepsilon, b]$ par une sous-famille finie $(]x_i, y'_i[)_{i\in\{0,1,\ldots,N_\varepsilon\}}$ de la famille des intervalles ouverts $(]x_i, y'_i[)_{i\in\mathbb{N}}$. Un raisonnement simple montre qu'alors

$$F(b) - F(a+\varepsilon) \leq \sum_{i=0}^{N_\varepsilon}(F(y'_i) - F(x_i)) \leq \sum_{i=0}^{\infty}(F(y'_i) - F(x_i)) \leq \sum_{i=0}^{\infty}(F(y_i) - F(x_i)) + 2\varepsilon.$$

En faisant tendre ε vers 0 on trouve

$$F(b) - F(a) \leq \sum_{i=0}^{\infty}(F(y_i) - F(x_i))$$

ce qui par définition de μ^* donne bien la minoration $\mu(]a, b]) \geq F(b) - F(a)$. $\square$

Cas des mesures de Radon. La formule

$$F(x) = \begin{cases} \mu(]0, x]) & \text{si } x \geq 0, \\ -\mu(]x, 0]) & \text{si } x < 0, \end{cases}$$

donne une correspondance bijective entre mesures de Radon μ sur $\mathbb{R}$ et fonctions $F : \mathbb{R} \longrightarrow \mathbb{R}$ croissantes continues à droite et nulles en 0. Ce résultat découle facilement du cas des mesures finies. On a encore l'égalité $\mu(]a, b]) = F(b) - F(a)$. Dans le cas particulier $F(x) = x$ la mesure μ est la mesure de Lebesgue.

3.6 Le théorème de représentation de Riesz

Soit X un espace métrique. On note $C_c(X)$ l'espace des fonctions continues à support compact sur X. Une forme linéaire J sur $C_c(X)$ est dite positive si $J(f) \geq 0$ dès que $f \geq 0$.

Si μ est une mesure de Radon sur X, on définit une forme linéaire J sur $C_c(X)$ en posant

$$J(f) = \int f\, d\mu.$$

Noter que l'intégrale est bien définie puisque $|f| \leq C\, 1_K$, où K est un compact de X, et μ est finie sur les compacts. De plus J est positive.

Le théorème de représentation de Riesz montre que sous des hypothèses convenables toute forme linéaire positive sur $C_c(X)$ est de ce type.

Théorème 3.6.1 *Soit X un espace métrique localement compact séparable, et soit J une forme linéaire positive sur $C_c(X)$. Il existe alors une unique mesure de Radon μ sur $(X, \mathcal{B}(X))$ telle que*

$$\forall f \in C_c(X),\ J(f) = \int f\, d\mu.$$

La mesure μ est régulière au sens où pour tout $A \in \mathcal{B}(X)$,

$$\begin{aligned}\lambda(A) &= \inf\{\lambda(U) : U \text{ ouvert }, A \subset U\} \\ &= \sup\{\lambda(F) : F \text{ compact }, F \subset A\}.\end{aligned}$$

De plus, pour tout ouvert U de X,

$$\mu(U) = \sup\{J(f) : f \in C_c(X),\ 0 \leq f \leq 1_U\}.$$

Exemple. Si $X = \mathbb{R}$, on peut prendre $J(f) = I(f)$, où $I(f)$ est comme ci-dessus l'intégrale de Riemann de la fonction f. On vérifie aisément que J est une forme linéaire positive sur $C_c(\mathbb{R})$. La mesure associée est (bien sûr) la mesure de Lebesgue. Cela fournit donc une autre construction de la mesure de Lebesgue (en supposant construite l'intégrale de Riemann des fonctions continues).

Nous ne donnons pas ici la preuve du Théorème 3.6.1 : voir le Théorème 10.1 de Briane et Pagès [2] ou le chapitre 2 de Rudin [7], qui donne un énoncé un peu plus précis.

Chapitre 4

Espaces L^p

Ce chapitre est consacré principalement à l'étude de l'espace L^p des fonctions dont la valeur absolue est de puissance p-ième intégrable. Les inégalités fondamentales de Hölder, de Minkowski et de Jensen constituent un outil important pour cette étude. On étudie notamment la structure d'espace de Banach de l'espace L^p, et dans le cas particulier $p = 2$ la structure d'espace de Hilbert de L^2. Les théorèmes de densité montrant qu'on peut approcher n'importe quelle fonction de L^p par des fonctions plus "régulières" jouent un rôle important dans beaucoup d'applications en analyse. En application de la structure hilbertienne de L^2, on établit le théorème de Radon-Nikodym, qui étant donné une mesure de référence permet de décomposer n'importe quelle autre mesure en la somme d'une mesure à densité par rapport à la mesure de référence et d'une mesure "étrangère".

4.1 Définition et inégalité de Hölder

Dans tout ce chapitre on considère un espace mesuré $(E, \mathcal{A}, \mu)$. Pour tout réel $p \geq 1$ on pose

$$\mathcal{L}^p(E, \mathcal{A}, \mu) = \{f : E \longrightarrow \mathbb{R} \text{ mesurable}; \int |f|^p d\mu < \infty\}$$

et on définit aussi

$$\mathcal{L}^\infty(E, \mathcal{A}, \mu) = \{f : E \longrightarrow \mathbb{R} \text{ mesurable}; \exists C \in \mathbb{R}_+ : |f| \leq C,\ \mu \text{ p.p.}\}.$$

On pourrait aussi considérer les espaces $\mathcal{L}^p_\mathbb{C}$ et $\mathcal{L}^\infty_\mathbb{C}$ obtenus en considérant des fonctions à valeurs complexes, mais dans ce chapitre nous nous intéresserons surtout au cas réel.

Pour chaque $p \in [1, \infty]$, on définit une relation d'équivalence sur $\mathcal{L}^p$ en posant

$$f \sim g \qquad \text{si et seulement si} \quad f = g,\ \mu \text{ p.p.}$$

Cela conduit à définir l'espace quotient

$$L^p(E, \mathcal{A}, \mu) = \mathcal{L}^p(E, \mathcal{A}, \mu)/ \sim \ .$$

Un élément de $L^p(E, \mathcal{A}, \mu)$ est donc une classe d'équivalence de fonctions égales μ p.p. Dans la suite on fera presque systématiquement l'abus d'écriture consistant à identifier un élement de $L^p(E, \mathcal{A}, \mu)$ à l'un de ses représentants.

Pour toute fonction $f : E \longrightarrow \mathbb{R}$ mesurable, on note pour $p \in [1, \infty[$,

$$\|f\|_p = \Big(\int |f|^p d\mu \Big)^{1/p}$$

(avec la convention $\infty^{1/p} = \infty$) et

$$\|f\|_\infty = \inf\{C \in [0, \infty] : |f| \leq C,\ \mu \text{ p.p.}\}$$

de façon que $\|f\| \leq \|f\|_\infty$, μ p.p. et que $\|f\|_\infty$ est le plus petit nombre dans $[0, \infty]$ avec cette propriété.

Soient $p, q \in [1, \infty]$. On dit que p et q sont des exposants conjugués si

$$\frac{1}{p} + \frac{1}{q} = 1.$$

En particulier, $p = 1$ et $q = \infty$ sont conjugués.

Théorème 4.1.1 (Inégalité de Hölder) *Soient p et q des exposants conjugués. Alors, si f et g sont deux fonctions mesurables de E dans $\mathbb{R}$,*

$$\int |fg|\, d\mu \leq \|f\|_p \|g\|_q\, .$$

En particulier, $fg \in L^1(E, \mathcal{A}, \mu)$ dès que $f \in L^p(E, \mathcal{A}, \mu)$ et $g \in L^q(E, \mathcal{A}, \mu)$.

Preuve. Si $\|f\|_p = 0$, on a $f = 0$, μ p.p., ce qui entraîne $\int |fg| d\mu = 0$, et l'inégalité est triviale. On peut donc supposer $\|f\|_p > 0$ et $\|g\|_q > 0$. Sans perte de généralité on peut aussi supposer $f \in L^p(E, \mathcal{A}, \mu)$ et $g \in L^q(E, \mathcal{A}, \mu)$.

Le cas $p = 1$, $q = \infty$ est facile : on a $|fg| \leq \|g\|_\infty |f|$, μ p.p., d'où

$$\int |fg|\, d\mu \leq \|g\|_\infty \int |f| d\mu = \|g\|_\infty \|f\|_1.$$

Supposons $1 < p < \infty$ (et donc $1 < q < \infty$).

Soit $\alpha \in]0, 1[$. On a pour tout $x \in \mathbb{R}_+$

$$x^\alpha - \alpha x \leq 1 - \alpha.$$

En effet la fonction $\varphi_\alpha(x) = x^\alpha - \alpha x$ a pour dérivée sur $]0, \infty[$, $\varphi'_\alpha(x) = \alpha(x^{\alpha-1} - 1)$ qui est positive sur $]0, 1[$ et négative sur $]1, \infty[$. Donc φ_α est maximale en $x = 1$, ce qui donne l'inégalité recherchée. En appliquant cette inégalité à $x = \frac{u}{v}$, où $u \geq 0$ et $v > 0$, on trouve

$$u^\alpha v^{1-\alpha} \leq \alpha u + (1 - \alpha) v,$$

inégalité qui reste vraie si $v = 0$. On prend alors $\alpha = \frac{1}{p}$ (donc $1 - \alpha = \frac{1}{q}$) puis

$$u = \frac{|f(x)|^p}{\|f\|_p^p} \qquad , \qquad v = \frac{|g(x)|^q}{\|g\|_q^q}$$

pour aboutir à

$$\frac{|f(x)g(x)|}{\|f\|_p\|g\|_q} \le \frac{1}{p}\frac{|f(x)|^p}{\|f\|_p^p} + \frac{1}{q}\frac{|g(x)|^q}{\|g\|_q^q}.$$

En intégrant cette dernière inégalité par rapport à μ, il vient

$$\frac{1}{\|f\|_p\|g\|_q}\int |fg| d\mu \le \frac{1}{p} + \frac{1}{q} = 1.$$

□

Exercice. Lorsque $1 < p < \infty$, montrer qu'il y a égalité dans l'inégalité de Hölder ssi il existe deux réels positifs α, β non tous deux nuls, tels que $\alpha |f|^p = \beta |g|^q$ μ p.p.

Le cas particulier $p = q = 2$ de l'inégalité de Hölder est l'inégalité de Cauchy-Schwarz

$$\int |fg|\, d\mu \le \Big(\int |f|^2 d\mu\Big)^{1/2}\Big(\int |g|^2 d\mu\Big)^{1/2}.$$

Considérons le cas particulier où μ est finie. En prenant $g = 1$, on trouve

$$\int |f|\, d\mu \le \mu(E)^{1/q}\, \|f\|_p$$

ce qui montre que $L^p \subset L^1$ pour tout $p \in]1,\infty]$. En remplaçant $|f|$ par $|f|^r$ $(r \ge 1)$ et en posant $r' = pr$, on trouve pour tous $1 \le r \le r' \le \infty$

$$\|f\|_r \le \mu(E)^{\frac{1}{r}-\frac{1}{r'}}\, \|f\|_{r'},$$

et donc $L^{r'} \subset L^r$ (toujours dans le cas où μ est finie). Lorsque μ est une mesure de probabilité on a $\|f\|_r \le \|f\|_{r'}$ pour tous $1 \le r \le r' \le \infty$.

Cette dernière inégalité peut être vue comme un cas particulier de l'inégalité de Jensen.

Théorème 4.1.2 (Inégalité de Jensen) *Supposons que μ est une mesure de probabilité, et soit $\varphi : \mathbb{R} \longrightarrow \mathbb{R}_+$ une fonction convexe. Alors, pour $f \in L^1(E, \mathcal{A}, \mu)$,*

$$\int \varphi \circ f\, d\mu \ge \varphi\Big(\int f\, d\mu\Big).$$

Remarque. L'intégrale $\int \varphi \circ f\, d\mu$ est bien définie comme intégrale d'une fonction mesurable positive.

Preuve. Soit

$$\mathcal{E}_\varphi = \{(a,b) \in \mathbb{R}^2 : \forall x \in \mathbb{R},\ \varphi(x) \ge ax + b\}.$$

Les propriétés bien connues des fonctions convexes assurent que

$$\forall x \in \mathbb{R}\ , \quad \varphi(x) = \sup_{(a,b)\in\mathcal{E}_\varphi} (ax + b).$$

En conséquence,

$$\begin{aligned}\int \varphi \circ f\, d\mu &\geq \sup_{(a,b)\in\mathcal{E}_\varphi} \int (af+b)d\mu \\ &= \sup_{(a,b)\in\mathcal{E}_\varphi} \Big(a\int f d\mu + b\Big) \\ &= \varphi\Big(\int f\, d\mu\Big)\end{aligned}$$

□

Exercice. Montrer que si $\mu(E) < \infty$ on a

$$\|f\|_\infty = \lim_{p\to\infty} \|f\|_p.$$

4.2 L'espace de Banach $L^p(E, \mathcal{A}, \mu)$

Théorème 4.2.1 (Inégalité de Minkowski) *Soit $p \in [1, \infty]$, et soient $f, g \in L^p(E, \mathcal{A}, \mu)$. Alors, $f + g \in L^p(E, \mathcal{A}, \mu)$ et*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p\,.$$

Preuve. Les cas $p = 1$ et $p = \infty$ sont faciles en utilisant simplement l'inégalité $|f + g| \leq |f| + |g|$. Supposons donc $1 < p < \infty$. En écrivant

$$|f+g|^p \leq 2^p(|f|^p + |g|^p)$$

on voit que $\int |f + g|^p d\mu < \infty$ et donc $f + g \in L^p$. Ensuite, en intégrant par rapport à μ l'inégalité

$$|f+g|^p \leq |f|\,|f+g|^{p-1} + |g|\,|f+g|^{p-1}$$

on trouve

$$\int |f+g|^p d\mu \leq \int |f|\,|f+g|^{p-1} d\mu + \int |g|\,|f+g|^{p-1} d\mu.$$

En appliquant l'inégalité de Hölder aux réels conjugués p et $q = p/(p-1)$, il vient

$$\int |f+g|^p d\mu \leq \|f\|_p \Big(\int |f+g|^p d\mu\Big)^{\frac{p-1}{p}} + \|g\|_p \Big(\int |f+g|^p d\mu\Big)^{\frac{p-1}{p}}.$$

Si $\int |f + g|^p d\mu = 0$, l'inégalité du théorème est triviale. Sinon on peut diviser chacun des deux membres de l'inégalité précédente par $(\int |f + g|^p d\mu)^{(p-1)/p}$ et on trouve le résultat recherché. □

Théorème 4.2.2 (Riesz) *Pour tout $p \in [1, \infty]$, l'espace $L^p(E, \mathcal{A}, \mu)$ muni de la norme $f \to \|f\|_p$ est un espace de Banach (i.e. un espace vectoriel normé complet).*

Preuve. On se limite au cas $1 \leq p < \infty$ (le cas $p = \infty$ est plus facile). Vérifions d'abord que $f \to \|f\|_p$ est une norme sur L^p. On a

$$\|f\|_p = 0 \Rightarrow \int |f|^p d\mu = 0 \Rightarrow f = 0 \ \mu \text{ p.p.}$$

ce qui signifie que $f = 0$ dans L^p (f appartient à la classe d'équivalence de 0). La propriété $\|\lambda f\|_p = |\lambda| \|f\|_p$ pour $\lambda \in \mathbb{R}$ est immédiate, et l'inégalité de Minkowski donne l'inégalité triangulaire.

Il reste à montrer que L^p muni de cette norme est complet. Soit $(f_n)_{n \geq 1}$ une suite de Cauchy dans L^p. Alors on peut choisir une suite d'entiers (k_n) strictement croissante de façon que pour tout $n \geq 1$,

$$\|f_{k_{n+1}} - f_{k_n}\|_p \leq 2^{-n}.$$

Posons $g_n = f_{k_n}$ et remarquons en utilisant le théorème de convergence monotone puis l'inégalité de Minkowski que

$$\begin{aligned}
\int \Big(\sum_{n=1}^{\infty} |g_{n+1} - g_n|\Big)^p d\mu &= \lim_{N \uparrow \infty} \uparrow \int \Big(\sum_{n=1}^{N} |g_{n+1} - g_n|\Big)^p d\mu \\
&\leq \lim_{N \uparrow \infty} \uparrow \Big(\sum_{n=1}^{N} \|g_{n+1} - g_n\|_p\Big)^p \\
&= \Big(\sum_{n=1}^{\infty} \|g_{n+1} - g_n\|_p\Big)^p \\
&< \infty.
\end{aligned}$$

On a donc

$$\sum_{n=1}^{\infty} |g_{n+1} - g_n| < \infty \ , \quad \mu \text{ p.p.}$$

et cela permet de poser

$$h = g_1 + \sum_{n=1}^{\infty} (g_{n+1} - g_n)$$

la série convergeant absolument sauf sur un ensemble de mesure nulle sur lequel on peut prendre une définition arbitraire de h (par exemple $h = 0$). La fonction h est alors mesurable. Puisque g_N converge vers h, μ p.p., on a $|h| = \liminf |g_N|$, μ p.p. et le lemme de Fatou montre immédiatement que

$$\int |h|^p d\mu \leq \liminf \int |g_N|^p d\mu \leq \sup_{N \geq 1} \int |g_N|^p d\mu < \infty,$$

puisque la suite f_n étant de Cauchy est bornée dans L^p. Enfin, à nouveau grâce au lemme de Fatou, on a

$$\|h - g_n\|_p^p = \int |h - g_n|^p d\mu \leq \liminf_{N \to \infty} \int |g_N - g_n|^p d\mu = \liminf_{N \to \infty} \|g_N - g_n\|_p^p \leq (2^{-n+1})^p$$

en majorant pour $N > n$, $\|g_N - g_n\|_p \leq \|g_{n+1} - g_n\|_p + \cdots + \|g_N - g_{N-1}\|_p \leq 2^{-n+1}$. L'inégalité précédente montre que g_n converge vers h dans L^p. Cela entraîne que f_n converge vers h et termine la preuve. □

Exemple. Si $E = \mathbb{N}$ et μ est la mesure de comptage, pour tout $p \in [1, \infty[$, l'espace L^p est l'espace des suites $a = (a_n)_{n \in \mathbb{N}}$ de réels tels que

$$\sum_{n=0}^{\infty} |a_n|^p < \infty$$

muni de la norme

$$\|a\|_p = \Big(\sum_{n=0}^{\infty} |a_n|^p \Big)^{1/p}.$$

L'espace L^∞ est simplement l'espace des suites $(a_n)_{n \in \mathbb{N}}$ qui sont bornées, muni de la norme $\|a\|_\infty = \sup(a_n)$. Remarquons que dans ce cas il n'y a pas d'ensemble non vide de mesure nulle et donc L^p coïncide avec $\mathcal{L}^p$. Cet espace est en général noté $\ell^p = \ell^p(\mathbb{N})$. Il joue un rôle important dans la théorie des espaces de Banach.

La dernière preuve fait apparaître un résultat intermédiaire qui mérite d'être énoncé.

Proposition 4.2.3 *Soit $p \in [1, \infty[$ et soit (f_n) une suite qui converge vers f dans $L^p(E, \mathcal{A}, \mu)$. Il existe alors une sous-suite (f_{k_n}) qui converge μ p.p. vers f.*

Remarque. Le résultat est aussi vrai pour $p = \infty$, mais dans ce cas l'extraction d'une sous-suite n'est pas nécessaire puisque la convergence L^∞ équivaut à une convergence uniforme sauf sur un ensemble de mesure nulle.

On peut se demander si inversement la convergence μ p.p. entraîne la convergence L^p. Cela n'est pas vrai, mais le théorème de convergence dominée montre que si :

(i) $f_n \longrightarrow f$, μ p.p.

(ii) Il existe une fonction $g \geq 0$ telle que $\int |f|^p d\mu < \infty$ et $\forall n$, $|f_n| \leq g$, μ p.p.

alors $f_n \longrightarrow f$ dans L^p.

Exercice. On suppose $\mu(E) < \infty$. Soit $p \in [1, \infty[$. Montrer que les conditions

(i) $f_n \longrightarrow f$, μ p.p.

(ii) Il existe $r > p$ tel que $\sup_n \displaystyle\int |f_n|^r d\mu < \infty$

entraînent $f_n \longrightarrow f$ dans L^p.

Le cas $p = 2$ du théorème de Riesz est particulièrement important puisque l'espace L^2 a une structure d'espace de Hilbert.

Théorème 4.2.4 *L'espace $L^2(E,\mathcal{A},\mu)$ muni du produit scalaire*

$$\langle f,g\rangle = \int fg\,d\mu$$

est un espace de Hilbert (réel).

Preuve. L'inégalité de Cauchy-Schwarz montre que si $f,g \in L^2$, fg est intégrable et donc $\langle f,g\rangle$ est bien défini. Ensuite il est clair que $(f,g) \longrightarrow \langle f,g\rangle$ définit une forme bilinéaire symétrique définie positive, et que la norme associée est la norme $\|f\|_2$. Le caractère complet découle du théorème de Riesz. □

On peut donc appliquer à $L^2(E,\mathcal{A},\mu)$ les résultats classiques sur les espaces de Hilbert. En particulier, si $\Phi : L^2(E,\mathcal{A},\mu) \longrightarrow \mathbb{R}$ est une forme linéaire continue, il existe un (unique) élément g de $L^2(E,\mathcal{A},\mu)$ tel que $\forall f \in L^2,\ \Phi(f) = \langle f,g\rangle$. Ce résultat nous sera utile dans la suite de ce chapitre.

Remarque. Comme les résultats précédents, le théorème ci-dessus s'étend au cas complexe. L'espace $L^2_{\mathbb{C}}(E,\mathcal{A},\mu)$ est un espace de Hilbert complexe pour le produit scalaire

$$\langle f,g\rangle = \int f\bar{g}\,d\mu.$$

4.3 Théorèmes de densité dans les espaces L^p

Si (E,d) est un espace métrique, une mesure μ sur $(E,\mathcal{B}(E))$ est dite extérieurement régulière si

$$\forall A \in \mathcal{B}(E)\ , \quad \mu(A) = \inf\{\mu(U) : U \text{ ouvert}, A \subset U\}.$$

Une fonction $f : E \longrightarrow \mathbb{R}$ est dite lipschitzienne s'il existe une constante K telle que

$$\forall x,y \in E\ , \quad |f(x) - f(y)| \leq K\,d(x,y).$$

Théorème 4.3.1 *Soit $p \in [1,\infty[$.*

(1) L'espace des fonctions étagées intégrables est dense dans $L^p(E,\mathcal{A},\mu)$.

(2) Si (E,d) est un espace métrique, et μ une mesure extérieurement régulière sur $(E,\mathcal{B}(E))$, l'espace des fonctions lipschitziennes bornées qui sont dans L^p est dense dans $L^p(E,\mathcal{B}(E),\mu)$.

(3) Si (E,d) est un espace métrique localement compact séparable, et μ une mesure de Radon sur E, alors l'espace des fonctions lipschitziennes à support compact est dense dans $L^p(E,\mathcal{B}(E),\mu)$.

Preuve. (1) En décomposant $f = f^+ - f^-$, il suffit de montrer que si $f \in L^p$ est positive, alors f est limite dans L^p d'une suite de fonctions étagées. On sait que

$$f = \lim_{n\to\infty} \uparrow \varphi_n$$

où pour chaque n, $0 \leq \varphi_n \leq f$ et φ_n est étagée. Alors, $\int |\varphi_n|^p d\mu \leq \int |f|^p d\mu < \infty$ et donc $\varphi_n \in L^p$ (ce qui pour une fonction étagée équivaut à $\varphi_n \in L^1$). Puisque $|f - \varphi_n|^p \leq f^p$, le théorème de convergence dominée donne

$$\lim_{n\to\infty} \int |f - \varphi_n|^p d\mu = 0.$$

(2) Il suffit de montrer que toute fonction étagée intégrable est limite dans L^p de fonctions lipschitziennes bornées. On se ramène aisément au cas $f = 1_A$, $A \in \mathcal{B}(E)$, $\mu(A) < \infty$. Soit alors $\varepsilon > 0$. On peut trouver un ouvert O contenant A tel que $\mu(O \backslash A) < (\varepsilon/2)^p$, et donc

$$\|1_O - 1_A\|_p < \frac{\varepsilon}{2}.$$

Ensuite, pour tout $k \geq 1$, on pose $\varphi_k(x) = (k\, d(x, O^c)) \wedge 1$. La fonction φ_k est lipschitzienne et $\varphi_k \uparrow 1_O$ quand $k \to \infty$. Par convergence dominée, $\int |1_O - \varphi_k|^p d\mu \longrightarrow 0$ quand $k \to \infty$, et donc on peut choisir k assez grand pour que

$$\|1_O - \varphi_k\|_p < \frac{\varepsilon}{2}.$$

Finalement,

$$\|1_A - \varphi_k\|_p \leq \|1_A - 1_O\|_p + \|1_O - \varphi_k\|_p < \varepsilon.$$

(3) On utilise le lemme suivant, dont la preuve est repoussée à la fin de la démonstration. Rappelons que si A est un sous-ensemble de E, $\overset{\circ}{A}$ désigne l'intérieur de A.

Lemme 4.3.2 *Soit E un espace métrique localement compact séparable. Alors il existe une suite croissante de compacts $(L_n)_{n\geq 1}$ tels que, pour tout n, $L_n \subset \overset{\circ}{L}_{n+1}$ et $E = \bigcup_{n\geq 1} L_n = \bigcup_{n\geq 1} \overset{\circ}{L}_n$.*

Il est facile de déduire du lemme que toute mesure de Radon μ sur E est extérieurement régulière (ce qui a déjà été vu, sans démonstration, dans l'énoncé du théorème de représentation de Riesz). En effet, si A est un borélien de E, on peut en considérant la restriction de μ à $\overset{\circ}{L}_n$ (qui est une mesure finie) appliquer un résultat de régularité extérieure du chapitre précédent et trouver pour chaque n un ouvert $O_n \subset \overset{\circ}{L}_n$ tel que $A \cap \overset{\circ}{L}_n \subset O_n$ et

$$\mu(O_n \backslash (A \cap \overset{\circ}{L}_n)) \leq \varepsilon\, 2^{-n}.$$

Alors la réunion O des O_n est un ouvert de E et

$$\mu(O \backslash A) \leq \sum_{n\geq 1} \mu(O_n \backslash (A \cap \overset{\circ}{L}_n)) \leq \varepsilon.$$

Ensuite, puisque μ est extérieurement régulière, on peut appliquer la partie (2) du théorème. On est ainsi ramené à montrer que toute fonction f lipschitzienne bornée telle

que $\int |f|^p d\mu < \infty$ est limite dans L^p de fonctions lipschitziennes à support compact (noter que celles-ci sont automatiquement dans L^p). Par convergence dominée, on a

$$\lim_{n\to\infty} \int_{(\overset{\circ}{L}_n)^c} |f|^p d\mu = 0,$$

et donc $\|f - f1_{\overset{\circ}{L}_n}\|_p \longrightarrow 0$. D'autre part, pour chaque n fixé, et pour tout $k \geq 1$, soit

$$\varphi_{n,k}(x) = k\, d(x, (\overset{\circ}{L}_n)^c) \wedge 1.$$

Alors $\varphi_{n,k} \in L^p$ puisque $\varphi_{n,k} \leq 1_{\overset{\circ}{L}_n}$. De plus, par convergence dominée à nouveau, on voit que pour chaque n fixé, $\varphi_{n,k}$ converge vers $1_{\overset{\circ}{L}_n}$ dans L^p quand $k \to \infty$. Finalement, en écrivant

$$\|f - f\varphi_{n,k}\|_p \leq \|f - f1_{\overset{\circ}{L}_n}\|_p + \|f1_{\overset{\circ}{L}_n} - f\varphi_{n,k}\|_p \leq \|f - f1_{\overset{\circ}{L}_n}\|_p + \|f\|_\infty \|1_{\overset{\circ}{L}_n} - \varphi_{n,k}\|_p$$

et en choisissant n puis k assez grands, on approche f dans L^p par la fonction $f\varphi_{n,k}$ qui est lipschitzienne à support compact. □

Preuve du lemme. On montre d'abord que E est réunion d'une suite croissante de compacts $(K_n)_{n\geq 1}$. Pour cela, soit $(x_p)_{p\geq 0}$ une suite dense dans E. Introduisons l'ensemble I de couples d'entiers défini par

$$I = \{(p,k) \in \mathbb{N}^2 : \bar{B}(x_p, 2^{-k}) \text{ est compact}\},$$

où $\bar{B}(x,r)$ désigne la boule fermée de centre x et de rayon r. En utilisant le fait que E est localement compact et la densité de la suite (x_p) il est facile de voir que

$$E = \bigcup_{(p,k)\in I} \bar{B}(x_p, 2^{-k}).$$

Par ailleurs, I étant dénombrable, on peut trouver une suite croissante de sous-ensembles finis I_n, $n \geq 1$ de I tels que I soit la réunion des I_n. Alors il suffit de poser

$$K_n = \bigcup_{(p,k)\in I_n} \bar{B}(x_p, 2^{-k})$$

pour avoir les propriétés recherchées.

Ensuite, on construit la suite (L_n) par récurrence sur n. On prend $L_1 = K_1$. Si on a construit L_n, on recouvre le compact $K_{n+1} \cup L_n$ par une réunion finie $V_1 \cup V_2 \cup \ldots \cup V_p$ de voisinages ouverts d'adhérence compacte de points de $K_{n+1} \cup L_n$, et on prend $L_{n+1} = \bar{V}_1 \cup \bar{V}_2 \cup \ldots \cup \bar{V}_p$. □

Conséquences. Pour $p \in [1, \infty[$, on a :

(i) L'espace $C_c(\mathbb{R}^d)$ des fonctions continues à support compact sur $\mathbb{R}^d$ est dense dans $L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$. On peut remplacer λ par n'importe quelle mesure de Radon sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

(ii) L'ensemble des fonctions en escalier (à support compact) est dense dans $L^p(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. En effet il sufit de vérifier que toute fonction $f \in C_c(\mathbb{R})$ est limite dans L^p de fonctions en escalier. Cela se voit en écrivant

$$f = \lim_{n\to\infty} \Big(\sum_{k\in\mathbb{Z}} f(\frac{k}{n})\, 1_{[\frac{k}{n},\frac{k+1}{n}[} \Big).$$

Application. Si $f \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$,

$$\hat{f}(\xi) \underset{|\xi|\to\infty}{\longrightarrow} 0.$$

On se ramène par densité au cas où f est une fonction en escalier : si f est limite dans L^1 d'une suite (φ_n) de fonctions en escalier,

$$\sup_{\xi\in\mathbb{R}} |\hat{f}(\xi) - \hat{\varphi}_n(\xi)| = \sup_{\xi\in\mathbb{R}} \Big| \int f(x)e^{ix\xi}dx - \int \varphi_n(x)e^{ix\xi}dx \Big| \le \|f - \varphi_n\|_1$$

qui tend vers 0 quand $n \to \infty$. Ensuite, si f est en escalier, $f = \sum_{j=1}^{p} \lambda_j\, 1_{]x_j,x_{j+1}[}$, on a

$$\hat{f}(\xi) = \sum_{j=1}^{p} \lambda_j \Big(\frac{e^{i\xi x_{j+1}} - e^{i\xi x_j}}{i\xi} \Big) \underset{|\xi|\to\infty}{\longrightarrow} 0,$$

d'où le résultat voulu.

4.4 Le théorème de Radon-Nikodym

Définition 4.4.1 *Soient μ et ν deux mesures sur $(E, \mathcal{A})$. On dit que:*

(i) *ν est absolument continue par rapport à μ (notation $\nu \ll \mu$) si*

$$\forall A \in \mathcal{A}, \quad \mu(A) = 0 \Rightarrow \nu(A) = 0.$$

(ii) *ν est étrangère à μ (notation $\nu \perp \mu$) s'il existe $N \in \mathcal{A}$ tel que $\mu(N) = 0$ et $\nu(N^c) = 0$.*

Exemple. Si f est mesurable positive , la mesure $\nu = f \cdot \mu$ définie par

$$\nu(A) = \int_A f\, d\mu$$

est absolument continue par rapport à μ.

Théorème 4.4.1 (Radon-Nikodym) *Soient μ et ν deux mesures σ-finies sur $(E, \mathcal{A})$. Il existe alors un unique couple (ν_a, ν_s) de mesures σ-finies sur $(E, \mathcal{A})$ telles que*

(1) $\nu = \nu_a + \nu_s$.

(2) $\nu_a \ll \mu$ *et* $\nu_s \perp \mu$.

De plus, il existe une fonction mesurable $g : E \longrightarrow \mathbb{R}_+$ *telle que*

$$\forall A \in \mathcal{A}, \quad \nu_a(A) = \int_A g \, d\mu$$

et la fonction g *est unique à un ensemble de* μ*-mesure nulle près.*

Preuve. On traite d'abord en détail le cas où les deux mesures μ et ν sont finies. L'extension au cas σ-fini ne présentera pas de difficulté.

Cas où $\mu \geq \nu$. Dans un premier temps, on suppose $\nu \leq \mu$, c'est-à-dire $\int g \, d\nu \leq \int g \, d\mu$ pour toute fonction mesurable positive g. Considérons alors l'application $\Phi : L^2(E, \mathcal{A}, \mu) \longrightarrow \mathbb{R}$ définie par

$$\Phi(f) = \int f \, d\nu.$$

Remarquons que l'intégrale a bien un sens puisque

$$\int |f| d\nu \leq \int |f| d\mu$$

et on sait que pour une mesure finie $L^2(\mu) \subset L^1(\mu)$. De plus, $\Phi(f)$ ne dépend pas du représentant de f choisi pour calculer $\int f d\nu$:

$$f = \tilde{f}, \; \mu \text{ p.p.} \Rightarrow f = \tilde{f}, \; \nu \text{ p.p.} \Rightarrow \int f d\nu = \int f d\tilde{\nu}.$$

L'inégalité de Cauchy-Schwarz montre que

$$|\Phi(f)| \leq \Big(\int f^2 d\nu \Big)^{1/2} \nu(E)^{1/2} \leq \Big(\int f^2 d\mu \Big)^{1/2} \nu(E)^{1/2} = \nu(E)^{1/2} \|f\|_{L^2(\mu)}.$$

Donc Φ est une forme linéaire continue sur $L^2(E, \mathcal{A}, \mu)$ et on sait alors qu'il existe une fonction $h \in L^2(E, \mathcal{A}, \mu)$ telle que

$$\forall f \in L^2(E, \mathcal{A}, \mu), \quad \Phi(f) = \langle f, h \rangle = \int f h \, d\mu.$$

En particulier, en prenant $f = 1_A$,

$$\forall A \in \mathcal{A}, \quad \nu(A) = \int_A h \, d\mu.$$

On peut aussi remarquer que $0 \leq h \leq 1$, μ p.p. En effet, pour tout $\varepsilon > 0$,

$$\mu(\{x : h(x) \geq 1{+}\varepsilon\}) \geq \nu(\{x : h(x) \geq 1{+}\varepsilon\}) = \int_{\{x : h(x) \geq 1+\varepsilon\}} h d\mu \geq (1{+}\varepsilon)\mu(\{x : h(x) \geq 1{+}\varepsilon\})$$

ce qui implique $\mu(\{x : h(x) \geq 1+\varepsilon\}) = 0$. On montre de même que $h \geq 0$ μ p.p. Remarquons que quitte à remplacer h par $(h \vee 0) \wedge 1$, on peut supposer $0 \leq h(x) \leq 1$ pour tout $x \in E$.

Cas général. On applique la première partie de la preuve aux mesures ν et $\mu+\nu$. Il existe donc une fonction mesurable h telle que $0 \leq h \leq 1$ et, pour toute fonction $f \in L^2(\mu+\nu)$,

$$\int f\,d\nu = \int fh\,d(\mu+\nu).$$

En particulier, pour toute fonction f mesurable bornée,

$$\int f\,d\nu = \int fh\,d\mu + \int fh\,d\nu$$

d'où

$$\int f(1-h)\,d\nu = \int fh\,d\mu.$$

En utilisant le théorème de convergence monotone, on voit que cette dernière égalité est vraie pour toute fonction f mesurable positive.

Posons $N = \{x \in E : h(x) = 1\}$. Alors en prenant $f = 1_N$, on voit que $\mu(N) = 0$. La mesure

$$\nu_s = 1_N \cdot \nu \qquad (\forall A \in \mathcal{A},\ \nu_s(A) = \nu(A \cap N))$$

est donc étrangère à μ. D'autre part, en remplaçant f par $1_{N^c}(1-h)^{-1}f$ dans l'égalité ci-dessus, on trouve que pour toute fonction f mesurable positive,

$$\int_{N^c} f\,d\nu = \int_{N^c} f\,\frac{h}{1-h}\,d\mu = \int fg\,d\mu,$$

où $g = 1_{N^c}\frac{h}{1-h}$. En posant

$$\nu_a = 1_{N^c} \cdot \nu = g \cdot \mu$$

on a bien les propriétés (1) et (2) du théorème, et la représentation annoncée pour ν_a.

L'unicité du couple (ν_a, ν_s) est facile. Si $(\tilde{\nu}_a, \tilde{\nu}_s)$ est un autre couple avec les propriétés (1) et (2), on a

$$\forall A \in \mathcal{A}, \quad \nu_a(A) - \tilde{\nu}_a(A) = \nu_s(A) - \tilde{\nu}_s(A).$$

Mais comme ν_s et $\tilde{\nu}_s$ sont portées respectivement par des ensembles N et $\tilde{N}$ de μ-mesure nulle, on a

$$\nu_s(A) - \tilde{\nu}_s(A) = \nu_s(A \cap (N \cup \tilde{N})) - \tilde{\nu}_s(A \cap (N \cup \tilde{N})) = \nu_a(A \cap (N \cup \tilde{N})) - \tilde{\nu}_a(A \cap (N \cup \tilde{N})) = 0$$

à cause de la propriété $\nu_a \ll \mu$, $\tilde{\nu}_a \ll \mu$. Enfin, pour obtenir l'unicité de g, on se donne une autre fonction $\tilde{g}$ avec la même propriété, et on observe que

$$\int_{\{\tilde{g}>g\}} \tilde{g}\,d\mu = \nu_a(\{\tilde{g} > g\}) = \int_{\{\tilde{g}>g\}} g\,d\mu,$$

d'où

$$\int_{\{\tilde{g}>g\}} (\tilde{g} - g)\,d\mu = 0$$

ce qui force $\tilde{g} \leq g$, μ p.p. et par symétrie $g = \tilde{g}$, μ p.p.

Il reste à s'affranchir de l'hypothèse supplémentaire que μ et ν sont finies. Si μ et ν sont seulement σ-finies, on peut construire une partition mesurable dénombrable $(E_n)_{n\in\mathbb{N}}$ de E de manière que $\mu(E_n) < \infty$ et $\nu(E_n) < \infty$ pour tout n. Notons μ_n la restriction de μ à E_n et ν_n la restriction de ν à E_n. En appliquant le début de la preuve on peut écrire pour tout $n \in \mathbb{N}$,

$$\nu_n = \nu_a^n + \nu_s^n$$

où $\nu_s^n \perp \mu_n$, et $\nu_a^n = g_n \cdot \mu_n$, la fonction mesurable g_n étant nulle sur E_n^c (puisque $\mu_n(E_n^c) = 0$, il est clair qu'on peut imposer cette dernière condition). On obtient le résultat du théorème en posant

$$\nu_a = \sum_{n\in\mathbb{N}} \nu_a^n \ , \quad \nu_s = \sum_{n\in\mathbb{N}} \nu_s^n \ , \quad g = \sum_{n\in\mathbb{N}} g_n.$$

(Dans la dernière somme, remarquer que pour chaque $x \in E$ il y a au plus une valeur de n pour laquelle $g_n(x) > 0$.) La vérification des propriétés d'unicité ne présente pas de difficulté.
$\square$

Chapitre 5

Mesures produits

Etant donné deux espaces mesurables munis chacun d'une mesure, on peut construire sur leur produit cartésien une mesure appelée la mesure produit. De plus l'intégrale d'une fonction définie sur l'espace produit peut être calculée en intégrant d'abord par rapport à la mesure sur le premier espace puis par rapport à la mesure sur le second, ou bien dans l'ordre inverse : c'est le fameux théorème de Fubini. Outre ses applications importantes en analyse (intégration par parties, convolution, etc.) ou en théorie des probabilités, le théorème de Fubini est un outil essentiel pour le calcul effectif des intégrales.

5.1 Généralités sur les espaces produits

Soient $(E, \mathcal{A})$ et $(F, \mathcal{B})$ deux espaces mesurables. On peut alors munir le produit $E \times F$ de la tribu-produit

$$\mathcal{A} \otimes \mathcal{B} = \sigma(A \times B\,;\, A \in \mathcal{A}, B \in \mathcal{B}).$$

Les ensembles de la forme $A \times B$ sont appelés pavés mesurables. Il est facile de vérifier que $\mathcal{A} \otimes \mathcal{B}$ est la plus petite tribu sur $E \times F$ qui rende mesurables les deux projections canoniques $\pi_1 : E \times F \longrightarrow E$ et $\pi_2 : E \times F \longrightarrow F$.

Soit $(G, \mathcal{C})$ un troisième espace mesurable, et soit $f : G \longrightarrow E \times F$. Notons $f(x) = (f_1(x), f_2(x))$. On a vu dans le Chapitre 1 que f est mesurable ($E \times F$ étant muni de la tribu produit) ssi les deux applications f_1 et f_2 le sont.

On étend facilement la définition de la tribu produit au cas d'un nombre fini quelconque d'espaces mesurables $(E_1, \mathcal{A}_1), \ldots, (E_n, \mathcal{A}_n)$:

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \cdots \otimes \mathcal{A}_n = \sigma(A_1 \times \cdots \times A_n\,;\, A_i \in \mathcal{A}_i)$$

et on a les propriétés d'"associativité" attendues, à savoir par exemple pour $n = 3$,

$$(\mathcal{A}_1 \otimes \mathcal{A}_2) \otimes \mathcal{A}_3 = \mathcal{A}_1 \otimes (\mathcal{A}_2 \otimes \mathcal{A}_3) = \mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \mathcal{A}_3\ .$$

Proposition 5.1.1 *Si E et F sont deux espaces métriques séparables, on a*

$$\mathcal{B}(E \times F) = \mathcal{B}(E) \otimes \mathcal{B}(F).$$

Preuve. L'inclusion $\mathcal{B}(E \times F) \supset \mathcal{B}(E) \otimes \mathcal{B}(F)$ est vraie sans hypothèse de séparabilité : elle découle de ce que les projections π_1 et π_2 sont continues donc mesurables pour la tribu $\mathcal{B}(E \times F)$.

Dans l'autre sens, on observe qu'on peut trouver un ensemble dénombrable d'ouverts $\mathcal{U} = \{U_n, n \geq 1\}$ de E tels que tout ouvert de E soit réunion d'une sous-famille de $\mathcal{U}$ (si (x_k) est une suite dense dans E, il suffit de prendre pour $\mathcal{U}$ les boules ouvertes de rayon rationnel centrées en l'un des x_k). Soit $\mathcal{V} = \{V_n, n \geq 1\}$ une famille analogue pour F. Pour tout ouvert O de $E \times F$ et tout $z = (x, y) \in O$, on sait que O contient un ouvert de la forme $U \times V$, où U, resp. V, est un ouvert de E, resp. de F, contenant x, resp. y. Il en découle que O doit être réunion (au plus dénombrable) d'une sous-famille de la famille $\{U_n \times V_m; n, m \geq 1\}$. Donc tout ouvert de $E \times F$ est mesurable pour $\mathcal{B}(E) \otimes \mathcal{B}(F)$ et cela entraîne $\mathcal{B}(E \times F) \subset \mathcal{B}(E) \otimes \mathcal{B}(F)$. □

On revient au cas où $(E, \mathcal{A})$ et $(F, \mathcal{B})$ sont deux espaces mesurables quelconques. Si $C \subset E \times F$, on pose pour $x \in E$

$$C_x = \{y \in F : (x, y) \in C\}$$

et pour $y \in F$,

$$C^y = \{x \in E : (x, y) \in C\}.$$

Si f est une fonction définie sur $E \times F$, on note pour $x \in E$, $f_x(y) = f(x, y)$ et pour $y \in F$, $f^y(x) = f(x, y)$.

Théorème 5.1.2 (i) *Soit $C \in \mathcal{A} \otimes \mathcal{B}$. Alors, pour tout $x \in E$, $C_x \in \mathcal{B}$ et pour tout $y \in F$, $C^y \in \mathcal{A}$.*
(ii) *Soit $f : E \times F \longrightarrow G$ une application mesurable pour la tribu produit $\mathcal{A} \otimes \mathcal{B}$. Alors, pour tout $x \in E$, f_x est $\mathcal{B}$-mesurable, et pour tout $y \in F$, f^y est $\mathcal{A}$-mesurable.*

Preuve. (i) Fixons $x \in E$ et posons

$$\mathcal{C} = \{C \in \mathcal{A} \otimes \mathcal{B} : C_x \in \mathcal{B}\}.$$

Alors $\mathcal{C}$ contient les pavés mesurables (si $C = A \times B$, $C_x = B$ ou $C_x = \varnothing$ selon que $x \in A$ ou $x \notin A$). Par ailleurs il est facile de vérifier que $\mathcal{C}$ est une tribu, et donc $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$.
(ii) Pour toute partie mesurable D de G,

$$f_x^{-1}(D) = \{y \in F : (x, y) \in f^{-1}(D)\} = (f^{-1}(D))_x$$

qui est dans $\mathcal{B}$ d'après (i). □

5.2 Construction de la mesure-produit

Théorème 5.2.1 *Soient μ et ν deux mesures σ-finies respectivement sur $(E, \mathcal{A})$ et sur $(F, \mathcal{B})$.*

(i) *Il existe une unique mesure m sur $(E \times F, \mathcal{A} \otimes \mathcal{B})$ telle que*

$$\forall A \in \mathcal{A},\ \forall B \in \mathcal{B},\ m(A \times B) = \mu(A)\nu(B)$$

(avec la convention usuelle $0 \cdot \infty = 0$). Cette mesure est σ-finie, et est notée $m = \mu \otimes \nu$.

(ii) *Pour tout $C \in \mathcal{A} \otimes \mathcal{B}$,*

$$\mu \otimes \nu(C) = \int_E \nu(C_x)\, \mu(dx) = \int_F \mu(C^y)\, \nu(dy).$$

Preuve. *Unicité*. Il existe une suite croissante $A_n \in \mathcal{A}$, resp. $B_n \in \mathcal{B}$, telle que $\mu(A_n) < \infty$, resp. $\mu(B_n) < \infty$, pour tout n, et $E = \cup A_n$, resp. $F = \cup B_n$. Alors, si $C_n = A_n \times B_n$, on a aussi

$$E \times F = \bigcup_n C_n.$$

Soient m et m' deux mesures sur $\mathcal{A} \otimes \mathcal{B}$ vérifiant la propriété énoncée en (i) du théorème. Alors,

- m et m' coïncident sur la classe des pavés mesurables, qui est stable par intersection finie et engendre la tribu $\mathcal{A} \otimes \mathcal{B}$;
- pour tout n, $m(C_n) = \mu(A_n)\nu(B_n) = m'(C_n) < \infty$.

D'après une conséquence du lemme de classe monotone vue dans le Chapitre 1, cela suffit pour dire que $m = m'$.

Existence. On pose pour tout $C \in \mathcal{A} \otimes \mathcal{B}$,

$$m(C) = \int_E \nu(C_x)\, \mu(dx). \tag{5.1}$$

Remarquons que $\nu(C_x)$ est bien définie pour tout $x \in E$ d'après le théorème précédent. Pour vérifier que la formule (5.1) a bien un sens il faut aussi montrer que l'application $x \longrightarrow \nu(C_x)$ est $\mathcal{A}$-mesurable.

Supposons d'abord ν finie et posons

$$\mathcal{G} = \{C \in \mathcal{A} \otimes \mathcal{B} : x \longrightarrow \nu(C_x) \text{ est } \mathcal{A}\text{-mesurable}\}.$$

Alors

- $\mathcal{G}$ contient les pavés mesurables : si $C = A \times B$, $\nu(C_x) = 1_A(x)\nu(B)$.
- $\mathcal{G}$ est une classe monotone : si $C \subset C'$, on a $\nu((C \backslash C')_x) = \nu(C_x) - \nu(C'_x)$ (parce que ν est finie !) et si C_n est une suite croissante, $\nu((\cup C_n)_x) = \lim \uparrow \nu((C_n)_x)$.

D'après le lemme de classe monotone, on a donc $\mathcal{G} = \mathcal{A} \otimes \mathcal{B}$, ce qui donne la mesurabilité recherchée pour l'application $x \longrightarrow \nu(C_x)$.

Dans le cas général où ν n'est pas finie mais seulement σ-finie, on choisit la suite (B_n) comme ci-dessus et on peut remplacer ν par $\nu_n(B) = \nu(B \cap B_n)$, pour obtenir que $x \longrightarrow \nu(C_x) = \lim \uparrow \nu_n(C_x)$ est mesurable pour tout $C \in \mathcal{A} \otimes \mathcal{B}$.

Il est ensuite facile de montrer que m est une mesure sur $\mathcal{A}\otimes\mathcal{B}$: si (C_n) est une famille de parties disjointes dans $\mathcal{A}\otimes\mathcal{B}$, les $(C_n)_x$ sont aussi disjoints pour tout $x\in E$, et donc

$$\begin{aligned} m\Big(\bigcup_n C_n\Big) &= \int_E \nu\Big(\bigcup_n (C_n)_x\Big)\,\mu(dx) \\ &= \int_E \sum_n \nu((C_n)_x)\,\mu(dx) \\ &= \sum_n \int_E \nu((C_n)_x)\,\mu(dx) \\ &= \sum_n m(C_n) \end{aligned}$$

l'interversion entre somme et intégrale étant justifiée par un résultat du Chapitre 2.

Il est immédiat que m vérifie la propriété

$$m(A\times B)=\mu(A)\nu(B).$$

Par ailleurs, si on définit m' par

$$m'(C)=\int_F \mu(C^y)\,\nu(dy),$$

les mêmes arguments montrent que m' est une mesure sur $\mathcal{A}\otimes\mathcal{B}$ qui vérifie la même propriété, ce qui d'après l'unicité entraîne $m=m'$. On en déduit l'assertion (ii) du théorème, ce qui complète la preuve. □

Remarques. (i) L'hypothèse de σ-finitude est essentielle au moins pour la partie (ii). En effet, si on prend $(E,\mathcal{A})=(F,\mathcal{B})=(\mathbb{R},\mathcal{B}(\mathbb{R}))$, $\mu=\lambda$ et ν la mesure de comptage, on remarque que pour $C=\{(x,x):x\in\mathbb{R}\}$,

$$\infty=\int \nu(C_x)\,\lambda(dx)\neq\int\lambda(C^y)\,\nu(dy)=0.$$

(ii) Si on a maintenant n mesures σ-finies $\mu_1,\ldots,\mu_n$, on peut définir le produit $\mu_1\otimes\cdots\otimes\mu_n$ en posant

$$\mu_1\otimes\cdots\otimes\mu_n=\mu_1\otimes(\mu_2\otimes(\cdots\otimes\mu_n)).$$

L'ordre des parenthèses n'a en fait pas d'importance car la mesure $\mu_1\otimes\cdots\otimes\mu_n$ est caractérisée par ses valeurs sur les pavés

$$\mu_1\otimes\cdots\otimes\mu_n(A_1\times\cdots\times A_n)=\mu_1(A_1)\ldots\mu_n(A_n).$$

Exemple. Si $(E,\mathcal{A})=(F,\mathcal{B})=(\mathbb{R},\mathcal{B}(\mathbb{R}))$, et $\mu=\nu=\lambda$, on vérifie facilement que $\lambda\otimes\lambda$ est la mesure de Lebesgue sur $\mathbb{R}^2$ (observer que la mesure de Lebesgue sur $\mathbb{R}^2$ est caractérisée par ses valeurs sur les rectangles $[a,b]\times[c,d]$, toujours d'après le lemme de classe monotone). Ceci se généralise en dimension supérieure et montre qu'il aurait suffi de construire la mesure de Lebesgue en dimension un.

5.3 Le théorème de Fubini

On commence par donner l'énoncé qui concerne les fonctions positives. Comme dans le paragraphe précédent, on considère deux espaces mesurables $(E, \mathcal{A})$ et $(F, \mathcal{B})$, et le produit $E \times F$ est muni de la tribu $\mathcal{A} \otimes \mathcal{B}$.

Théorème 5.3.1 (Fubini-Tonnelli) *Soient μ et ν deux mesures σ-finies respectivement sur $(E, \mathcal{A})$ et sur $(F, \mathcal{B})$, et soit $f : E \times F \longrightarrow [0, \infty]$ une fonction mesurable.*
(i) *Les fonctions*

$$x \longrightarrow \int f(x,y)\,\nu(dy)$$
$$y \longrightarrow \int f(x,y)\,\mu(dx)$$

sont respectivement $\mathcal{A}$-mesurable et $\mathcal{B}$-mesurable.
(ii) *On a*

$$\int_{E\times F} f\,d\mu\otimes\nu = \int_E \Big(\int_F f(x,y)\,\nu(dy)\Big)\mu(dx) = \int_F \Big(\int_E f(x,y)\,\mu(dx)\Big)\nu(dy).$$

Preuve. (i) Soit $C \in \mathcal{A}\otimes\mathcal{B}$. Si $f = 1_C$, on a déjà vu que la fonction $x \longrightarrow \int f(x,y)\nu(dy) = \nu(C_x)$ est $\mathcal{A}$-mesurable, et de même $y \longrightarrow \int f(x,y)\mu(dx) = \mu(C^y)$ est $\mathcal{B}$-mesurable. Par linéarité, on en déduit que le résultat de (i) est vrai pour toute fonction étagée positive. Enfin, si f est quelconque, on peut écrire $f = \lim \uparrow f_n$, où les fonctions f_n sont étagées positives, et on utilise le fait qu'alors

$$\int f(x,y)\,\nu(dy) = \lim\uparrow \int f_n(x,y)\,\nu(dy)$$

et de même pour $\int f(x,y)\,\mu(dx)$.

(ii) Pour $f = 1_C$, l'égalité annoncée est

$$\mu\otimes\nu(C) = \int_E \nu(C_x)\,\mu(dx) = \int_F \mu(C^x)\,\nu(dy)$$

et a déjà été vue dans le paragraphe précédent. On en déduit par linéarité le résultat voulu quand f est étagée positive, puis par limite croissante pour f quelconque : on remarque par exemple que si $f = \lim\uparrow f_n$,

$$\int_E \Big(\int_F f(x,y)\,\nu(dy)\Big)\mu(dx) = \lim\uparrow \int_E \Big(\int_F f_n(x,y)\,\nu(dy)\Big)\mu(dx)$$

par une double application du théorème de convergence monotone. □

Nous passons maintenant au cas de fonctions de signe quelconque. On conserve les hypothèses du théorème précédent.

Théorème 5.3.2 (Fubini-Lebesgue) *Soit* $f \in \mathcal{L}^1(E \times F, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$ *(ou* $f \in \mathcal{L}^1_{\mathbb{C}}(E \times F, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$*). Alors*

(a) $\mu(dx)$ *p.p. la fonction* $y \longrightarrow f(x,y)$ *est dans* $\mathcal{L}^1(F, \mathcal{B}, \nu)$*,*

$\nu(dy)$ *p.p. la fonction* $x \longrightarrow f(x,y)$ *est dans* $\mathcal{L}^1(E, \mathcal{A}, \mu)$*.*

(b) *Les fonctions* $x \longrightarrow \int f(x,y)\,\nu(dy)$ *et* $y \longrightarrow \int f(x,y)\,\mu(dx)$*, bien définies sauf sur un ensemble mesurable de mesure nulle, sont respectivement dans* $\mathcal{L}^1(E, \mathcal{A}, \mu)$ *et* $\mathcal{L}^1(F, \mathcal{B}, \nu)$*.*

(c) *On a*

$$\int_{E\times F} f\, d\mu \otimes \nu = \int_E \Big(\int_F f(x,y)\,\nu(dy) \Big) \mu(dx) = \int_F \Big(\int_E f(x,y)\,\mu(dx) \Big) \nu(dy).$$

Preuve. (a) En appliquant le théorème précédent à $|f|$,

$$\int_E \Big(\int_F |f(x,y)|\,\nu(dy) \Big) \mu(dx) = \int |f|\, d\mu \otimes \nu < \infty.$$

cela entraîne que $\mu(dx)$ p.p.

$$\int_F |f(x,y)|\,\nu(dy) < \infty$$

et donc la fonction $y \longrightarrow f(x,y)$, dont on sait déjà qu'elle est mesurable, est dans $\mathcal{L}^1(F, \mathcal{B}, \nu)$.

(b) En écrivant $f = f^+ - f^-$ et en utilisant le théorème précédent, on voit que

$$x \longrightarrow \int f(x,y)\,\nu(dy) = \int f^+(x,y)\,\nu(dy) - \int f^-(x,y)\,\nu(dy)$$

est mesurable (pour être précis, il faudrait donner une valeur arbitraire, par exemple 0, à l'intégrale $\int f(x,y)\,\nu(dy)$ pour les x tels que $\int |f(x,y)|\,\nu(dy) = \infty$, qui forment un ensemble de mesure nulle). De plus,

$$\int_E \Big| \int_F f(x,y)\,\nu(dy) \Big| \mu(dx) \le \int_E \Big(\int_F |f(x,y)|\,\nu(dy) \Big) \mu(dx) = \int |f|\, d\mu \otimes \nu < \infty.$$

(c) Il suffit de faire la différence terme à terme dans les égalités

$$\int_E \Big(\int_F f^+(x,y)\,\nu(dy) \Big) \mu(dx) = \int_{E\times F} f^+\, d\mu \otimes \nu$$
$$\int_E \Big(\int_F f^-(x,y)\,\nu(dy) \Big) \mu(dx) = \int_{E\times F} f^-\, d\mu \otimes \nu.$$

□

Remarque. L'hypothèse $f \in \mathcal{L}^1(\mu \otimes \nu)$ est cruciale. Il peut arriver en effet que les propriétés (a) et (b) soient toutes les deux satisfaites, et donc que les quantités

$$\int_E \Big(\int_F f(x,y)\,\nu(dy) \Big) \mu(dx) \text{ et } \int_F \Big(\int_E f(x,y)\,\mu(dx) \Big) \nu(dy)$$

soient bien définies, sans que ces quantités soient égales. Pour donner un exemple, considérons la fonction

$$f(x,y) = 2e^{-2xy} - e^{-xy}$$

définie pour $(x,y) \in]0,\infty[\times]0,1]$. Alors, pour tout $y \in]0,1]$,

$$\int_{]0,\infty[} f(x,y)\,dx = 2\int_0^\infty e^{-2xy}dx - \int_0^\infty e^{-xy}dx = 0$$

et pour tout $x > 0$,

$$\int_{]0,1]} f(x,y)dy = 2\int_0^1 e^{-2xy}dy - \int_0^1 e^{-xy}dy = \frac{e^{-x} - e^{-2x}}{x}.$$

On voit alors que

$$\int_{]0,1]} \Big(\int_{]0,\infty[} f(x,y)\,dx \Big) dy = 0$$

alors que

$$\int_{]0,\infty[} \Big(\int_{]0,1]} f(x,y)dy \Big) dx = \int_0^\infty \frac{e^{-x} - e^{-2x}}{x} dx > 0.$$

Evidemment dans cet exemple on a

$$\int_{]0,\infty[\times]0,1[} |f(x,y)|\,dxdy = \infty.$$

En pratique, il faut se souvenir que l'application du théorème de Fubini est toujours justifiée pour des fonctions mesurables positives, et que dans le cas de fonctions de signe quelconque, il faut s'assurer que

$$\int |f|\,d\mu\otimes\nu < \infty$$

ce qui se fait le plus souvent en appliquant le cas des fonctions positives.

Notation. Lorsque l'application du théorème de Fubini est justifiée (et seulement dans ce cas), on omet souvent les parenthèses et on écrit

$$\int f\,d\mu\otimes\nu = \int_E \int_F f(x,y)\,\mu(dx)\nu(dy).$$

5.4 Applications

5.4.1 Intégration par parties

Soient f et g deux fonctions mesurables de $\mathbb{R}$ dans $\mathbb{R}$ localement intégrables (i.e. intégrables sur tout compact pour la mesure de Lebesgue). On pose pour $x \in \mathbb{R}$,

$$F(x) = \int_0^x f(t)\,dt \quad \left(= \begin{cases} \int_{[0,x]} f(t)\,dt & \text{si } x \geq 0 \\ -\int_{[x,0]} f(t)\,dt & \text{si } x < 0 \end{cases} \right)$$

$$G(x) = \int_0^x g(t)\,dt.$$

Alors, pour tous $a < b$,

$$F(b)G(b) = F(a)G(a) + \int_a^b f(t)G(t)dt + \int_a^b F(t)g(t)dt.$$

On voit facilement que cette égalité équivaut à

$$\int_a^b f(t)(G(t) - G(a))\,dt = \int_a^b (F(b) - F(t))g(t)\,dt.$$

Pour établir cette dernière égalité, on écrit

$$\begin{aligned}
\int_a^b f(t)(G(t) - G(a))\,dt &= \int_a^b f(t)\Big(\int_a^t g(s)ds\Big)dt \\
&= \int_a^b \Big(\int_a^b 1_{\{s\le t\}} f(t)g(s)ds\Big)dt \\
&= \int_a^b \Big(\int_a^b 1_{\{s\le t\}} f(t)g(s)dt\Big)ds \\
&= \int_a^b g(s)\Big(\int_s^b f(t)dt\Big)ds \\
&= \int_a^b g(s)(F(b) - F(s))ds.
\end{aligned}$$

Dans la troisième égalité on a appliqué le théorème de Fubini-Lebesgue à la fonction

$$\varphi(s,t) = 1_{\{s\le t\}} f(t)g(s)$$

en observant que, grâce au théorème de Fubini-Tonnelli,

$$\int_{[a,b]^2} |\varphi(s,t)|dsdt \le \int_{[a,b]^2} |f(t)||g(s)|dsdt = \Big(\int_{[a,b]} |f(t)|dt\Big)\Big(\int_{[a,b]} |g(s)|ds\Big) < \infty.$$

5.4.2 Convolution

Si f et g sont deux fonctions mesurables sur $\mathbb{R}^d$, la convolution

$$f * g(x) = \int_{\mathbb{R}^d} f(x-y)g(y)\,dy$$

est bien définie à condition que

$$\int_{\mathbb{R}^d} |f(x-y)g(y)|\,dy < \infty.$$

Dans ce cas, l'invariance de la mesure de Lebesgue par translation et par la symétrie $y \to -y$ entraîne aussitôt que $g * f(x)$ est bien définie et $g * f(x) = f * g(x)$.

Proposition 5.4.1 *Soient $f, g \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$. Alors, pour λ presque tout $x \in \mathbb{R}^d$, la convolution $f * g(x)$ est bien définie. De plus, $f * g \in L^1(\lambda)$ et $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$.*

Remarque. Cela a bien un sens de dire qu'une fonction définie λ presque partout est dans $L^1(\lambda)$: on peut choisir de manière arbitraire le prolongement sur l'ensemble où la fonction n'est pas définie.

Preuve. D'après le théorème de Fubini-Tonnelli,

$$\begin{aligned}
\int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |f(x-t)||g(t)| dt \Big) dx &= \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |f(x-t)||g(t)| dx \Big) dt \\
&= \int_{\mathbb{R}^d} |g(t)| \Big(\int_{\mathbb{R}^d} |f(x-t)| dx \Big) dt \\
&= \Big(\int_{\mathbb{R}^d} |g(t)| dt \Big) \Big(\int_{\mathbb{R}^d} |f(x)| dx \Big) \\
&< \infty
\end{aligned}$$

ce qui montre que

$$\int_{\mathbb{R}^d} |f(x-t)||g(t)| dt < \infty \qquad dx \text{ p.p.}$$

et donne la première assertion. Pour la seconde, on utilise encore le calcul précédent pour écrire

$$\int_{\mathbb{R}^d} |f * g(x)| dx \leq \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |f(x-t)||g(t)| dt \Big) dx = \|f\|_1 \|g\|_1 < \infty.$$

□

La proposition suivante donne un autre cadre dans lequel on peut considérer la convolution de f et g.

Proposition 5.4.2 *Soit $p \in [1, \infty[$, et soit $q \in]1, \infty]$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Soient $f \in L^p((\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$ et $g \in L^q(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$. Alors, pour tout $x \in \mathbb{R}^d$, la convolution $f * g(x)$ est bien définie et $f * g$ est uniformément continue et bornée sur $\mathbb{R}^d$.*

Preuve. L'inégalité de Hölder donne

$$\int_{\mathbb{R}^d} |f(x-y) g(y)| \, dy \leq \Big(\int |f(x-y)|^p dy \Big)^{1/p} \|g\|_q = \|f\|_p \|g\|_q.$$

Cela donne la première assertion et montre aussi que $f * g$ est bornée par $\|f\|_p \|g\|_q$. Pour l'uniforme continuité, on utilise le lemme suivant.

Lemme 5.4.3 *Notons $\sigma_x(y) = y - x$. Pour $f \in L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$, $p \in [1, \infty[$, l'application $x \longrightarrow f \circ \sigma_x$ est uniformément continue de $\mathbb{R}^d$ dans $L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$.*

Si on admet le lemme, il est facile de compléter la preuve de la proposition : pour $x, x' \in \mathbb{R}^d$,

$$\begin{aligned} |f * g(x) - f * g(x')| &\leq \int |f(x-y) - f(x'-y)||g(y)|\, dy \\ &\leq \|g\|_q \Big(\int |f(x-y) - f(x'-y)|^p dy \Big)^{1/p} \\ &= \|g\|_q \, \|f \circ \sigma_{-x} - f \circ \sigma_{-x'}\|_p \end{aligned}$$

et on utilise le lemme pour dire que $\|f \circ \sigma_{-x} - f \circ \sigma_{-x'}\|_p$ tend vers 0 quand $x - x'$ tend vers 0.

Preuve du lemme. Supposons d'abord $f \in C_c(\mathbb{R}^d)$. Alors,

$$\int |f \circ \sigma_x - f \circ \sigma_y|^p d\lambda = \int |f(z-x) - f(z-y)|^p dz = \int |f(z) - f(z-(y-x))|^p dz$$

qui tend vers 0 quand $y - x \to 0$ par convergence dominée. Dans le cas général, on peut trouver une suite $f_n \in C_c(\mathbb{R}^d)$ qui converge vers f dans $L^p(\lambda)$ (cf Chapitre 4). Alors

$$\begin{aligned} \|f \circ \sigma_x - f \circ \sigma_y\|_p &\leq \|f \circ \sigma_x - f_n \circ \sigma_x\|_p + \|f_n \circ \sigma_x - f_n \circ \sigma_y\|_p + \|f_n \circ \sigma_y - f \circ \sigma_y\|_p \\ &= 2\|f - f_n\|_p + \|f_n \circ \sigma_x - f_n \circ \sigma_y\|_p. \end{aligned}$$

Pour $\varepsilon > 0$, on choisit d'abord n tel que $\|f - f_n\|_p < \varepsilon/4$, puis $\delta > 0$ tel que $\|f_n \circ \sigma_x - f_n \circ \sigma_y\|_p \leq \varepsilon/2$ si $|x-y| < \delta$. Les inégalités précédentes montrent alors que $\|f \circ \sigma_x - f \circ \sigma_y\|_p \leq \varepsilon$ si $|x - y| < \delta$. □

Approximations de la mesure de Dirac. On dit qu'une suite φ_n dans $C_c(\mathbb{R}^d)$ est une approximation de δ_0 si :

- Il existe un compact K tel que $\mathrm{supp}(\varphi_n) \subset K$ pour tout n.
- Pour tout n, $\varphi_n \geq 0$ et
$$\int_{\mathbb{R}^d} \varphi_n(x)\, dx = 1.$$
- Pour tout $\delta > 0$,
$$\lim_{n\to\infty} \int_{\{|x|>\delta\}} \varphi_n(x)\, dx = 0.$$

Il est facile de construire des approximations de δ_0. Si $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}_+$ est une fonction continue à support compact telle que $\int \varphi(x) dx = 1$, il suffit de poser

$$\varphi_n(x) = n^d \varphi(nx)\ , \quad x \in \mathbb{R}^d.$$

On peut même s'arranger pour que les fonctions φ_n soient de classe C^∞ : prendre par exemple

$$\varphi(x) = c \exp\Big(- \frac{1}{1 - |x|^2} \Big)\, 1_{\{|x|<1\}},$$

la constante $c > 0$ étant choisie pour que la condition $\int \varphi(x) dx = 1$ soit satisfaite.

Proposition 5.4.4 *Soit* (φ_n) *une approximation de* δ_0.
(i) *Si* $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ *est continue, on a* $\varphi_n * f \longrightarrow f$ *quand* $n \to \infty$*, uniformément sur tout compact.*
(ii) *Si* $f \in L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$*, avec* $p \in [1, \infty[$*, on a* $\varphi_n * f \longrightarrow f$ *dans* L^p.

Preuve. La partie (i) est facile à établir, en écrivant

$$\varphi_n * f(x) = \int_{|y|\le\delta} f(x-y)\varphi_n(y)dy + \int_{|y|>\delta} f(x-y)\varphi_n(y)dy$$

et en utilisant la continuité de f. Pour la partie (ii), on observe que si $f, g \in L^p(\mathbb{R}^d, \lambda)$,

$$\begin{aligned}\int |\varphi_n * f(x) - \varphi_n * g(x)|^p dx &\le \int \Big(\int \varphi_n(x-y)|f(y)-g(y)|dy \Big)^p dx \\ &\le \int \Big(\int \varphi_n(x-y)|f(y)-g(y)|^p dy \Big) dx \\ &= \int |f(y)-g(y)|^p \Big(\int \varphi_n(x-y)dx \Big) dy \\ &= \int |f(y)-g(y)|^p dy\end{aligned}$$

où la deuxième inégalité est une conséquence de l'inégalité de Jensen (observer que $\varphi_n(x-y)dy$ est une mesure de probabilité). Cette majoration permet de se ramener au cas où $f \in C_c(\mathbb{R}^d)$, et alors le résultat découle de (i) et du théorème de convergence dominée. □

Application. En dimension $d = 1$, on peut prendre

$$\varphi_n(x) = c_n(1-x^2)^n \, 1_{\{|x|\le 1\}}$$

où la constante c_n est choisie pour que $\int \varphi_n(x)dx = 1$. Soit alors $[a, b]$ un intervalle contenu dans $]0, 1[$, et soit f une fonction continue sur $[a, b]$. On peut facilement prolonger f en une fonction continue sur $\mathbb{R}$ et à support compact contenu dans $[0, 1]$ (prendre par exemple f affine sur les intervalles $[0, a]$ et $[b, 1]$. Alors,

$$\varphi_n * f(x) = c_n \int (1-(x-y)^2)^n 1_{\{|x-y|\le 1\}} f(y)dy \longrightarrow f(x)$$

uniformément sur $[a, b]$. Pour $x \in [a, b]$, on peut clairement enlever l'indicatrice $1_{\{|x-y|\le 1\}}$, et on voit que f est limite uniforme sur $[a, b]$ de polynômes (théorème de Stone-Weierstrass).

5.4.3 Calcul du volume de la boule unité

On note ici B_d la boule unité fermée de $\mathbb{R}^d$, et λ_d la mesure de Lebesgue sur $\mathbb{R}^d$. En vue de calculer $\gamma_d = \lambda_d(B_d)$ on observe d'abord que pour tout $a > 0$, l'image de λ_d par l'application $x \longrightarrow ax$ est $a^{-d}\lambda_d$: pour tout $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda_d(a^{-1}A) = a^{-d}\lambda_d(A)$$

(il suffit de le vérifier lorsque A est un pavé, et alors c'est évident). En particulier,

$$\lambda_d(aB_d) = a^d \lambda_d(B_d).$$

Ensuite on écrit en utilisant le théorème de Fubini, si $d \geq 2$,

$$\begin{aligned}
\gamma_d = \int_{\mathbb{R}^d} 1_{B_d}(x)dx &= \int_{\mathbb{R}^d} 1_{\{x_1^2+\cdots+x_d^2\leq 1\}} dx_1 \dots dx_d \\
&= \int_{-1}^{1} \Big(\int_{\mathbb{R}^{d-1}} 1_{\{x_1^2+\cdots+x_{d-1}^2\leq 1-x_d^2\}} dx_1 \dots dx_{d-1} \Big) dx_d \\
&= \int_{-1}^{1} \lambda_{d-1}\Big(\sqrt{1-x_d^2} B_{d-1}\Big)\, dx_d \\
&= \gamma_{d-1} \int_{-1}^{1} (1-x_d^2)^{(d-1)/2}\, dx_d \\
&= \gamma_{d-1} I_{d-1}
\end{aligned}$$

à condition de poser pour tout entier $n \geq 0$,

$$I_n = \int_{-1}^{1} (1-x^2)^{n/2} dx.$$

Une intégration par parties simple montre que pour $n \geq 2$,

$$I_n = \frac{n}{n+1} I_{n-2}.$$

En utilisant les cas particuliers $I_0 = 2$, $I_1 = \pi/2$, on en déduit par récurrence que pour tout $d \geq 2$,

$$I_{d-1} I_{d-2} = \frac{2\pi}{d}.$$

En conséquence, pour $d \geq 3$,

$$\gamma_d = I_{d-1} I_{d-2} \gamma_{d-2} = \frac{2\pi}{d}\, \gamma_{d-2}.$$

A partir des cas particuliers $\gamma_1 = 2$, $\gamma_2 = \gamma_1 I_1 = \pi$, on en déduit

$$\gamma_{2k} = \frac{\pi^k}{k!}\ , \qquad \gamma_{2k+1} = \frac{\pi^k}{(k+\frac{1}{2})(k-\frac{1}{2})\cdots\frac{3}{2}\cdot\frac{1}{2}}$$

ce qu'on peut regrouper dans la formule

$$\gamma_d = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}.$$

Chapitre 6

Mesures signées

A la différence des chapitres précédents, on considère ici des mesures signées, pouvant prendre aussi bien des valeurs négatives que des valeurs positives. Le résultat principal de ce chapitre est la décomposition de Jordan, qui fournit une écriture minimale d'une telle mesure signée comme la différence de deux mesures positives portées par des ensembles mesurables disjoints. A titre d'application, on établit un théorème important d'analyse fonctionnelle, qui affirme que pour deux exposants p et q conjugués ($\frac{1}{p} + \frac{1}{q} = 1$) l'espace L^q est le dual topologique de L^p.

6.1 Définition et variation totale

Définition 6.1.1 *Soit $(E, \mathcal{A})$ un espace mesurable. Une mesure signée μ sur $(E, \mathcal{A})$ est une application $\mu : \mathcal{A} \longrightarrow \mathbb{R}$ telle que $\mu(\varnothing) = 0$ et que pour toute famille $(A_n)_{n\in\mathbb{N}}$ d'éléments disjoints de $\mathcal{A}$, la série*

$$\sum_{n\in\mathbb{N}} \mu(A_n)$$

converge absolument, et

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

Théorème 6.1.1 *Soit μ une mesure signée sur $(E, \mathcal{A})$. Pour tout $A \in \mathcal{A}$, posons*

$$|\mu|(A) = \sup\Big(\sum_{n\in\mathbb{N}} |\mu(A_n)| : A = \bigcup_{n\in\mathbb{N}} A_n\,,\ A_n \text{ disjoints}\Big)$$

où le supremum porte sur toutes les écritures de A comme réunion d'une famille dénombrable $(A_n)_{n\in\mathbb{N}}$ de parties mesurables disjointes. Alors $|\mu|$ est une mesure positive finie sur $(E, \mathcal{A})$, et pour tout $A \in \mathcal{A}$, $|\mu(A)| \le |\mu|(A)$.

Preuve. On montre d'abord que $|\mu|$ est une mesure positive. Soit $(B_i)_{i\in\mathbb{N}}$ une famille de parties mesurables disjointes, et $B = \bigcup_{i\in\mathbb{N}} B_i$. Par définition, si $t_i \in [0, |\mu|(B_i)[$ (ou $t_i = 0$

dans le cas $|\mu|(B_i) = 0$), on peut trouver une partition[1] mesurable $B_i = \bigcup_{n\in\mathbb{N}} A_{n,i}$, de façon que

$$\sum_{n\in\mathbb{N}} |\mu(A_{n,i})| \geq t_i.$$

Alors $(A_{n,i})_{n,i\in\mathbb{N}}$ est une partition dénombrable de B, et donc

$$|\mu|(B) \geq \sum_{i\in\mathbb{N}} \sum_{n\in\mathbb{N}} |\mu(A_{n,i}| \geq \sum_{i\in\mathbb{N}} t_i.$$

Puisque les t_i peuvent être choisis arbitrairement proches des $|\mu|(B_i)$, il en découle que

$$|\mu|(B) \geq \sum_{i\in\mathbb{N}} |\mu|(B_i).$$

Pour obtenir l'inégalité inverse, soit $(A_n)_{n\in\mathbb{N}}$ une partition de B. Alors

$$\begin{aligned}
\sum_{n\in\mathbb{N}} |\mu(A_n)| &= \sum_{n\in\mathbb{N}} |\sum_{i\in\mathbb{N}} \mu(A_n \cap B_i)| \\
&\leq \sum_{n\in\mathbb{N}} \sum_{i\in\mathbb{N}} |\mu(A_n \cap B_i)| \\
&= \sum_{i\in\mathbb{N}} \sum_{n\in\mathbb{N}} |\mu(A_n \cap B_i)| \\
&\leq \sum_{i\in\mathbb{N}} |\mu|(B_i),
\end{aligned}$$

la dernière inégalité découlant du fait que les $A_n \cap B_i$, $n \in \mathbb{N}$ forment une partition de B_i, et de la définition de $|\mu|(B_i)$. En prenant le supremum sur les partitions $(A_n)_{n\in\mathbb{N}}$ de B, on trouve

$$|\mu|(B) \leq \sum_{i\in\mathbb{N}} |\mu|(B_i)$$

ce qui achève de montrer que $|\mu|$ est une mesure positive.

Comme l'inégalité $|\mu(A)| \leq |\mu|(A)$ est immédiate, il reste à établir que $|\mu|$ est une mesure finie.

Lemme 6.1.2 *Si $A \in \mathcal{A}$ est tel que $|\mu|(A) = \infty$, alors il existe deux parties mesurables disjointes B et C telles que $A = B \cup C$ et $|\mu(B)| > 1$, $|\mu|(C) = \infty$.*

Preuve du lemme. Puisque $|\mu|(A) = \infty$, on peut trouver une partition mesurable $A = \bigcup_{n\in\mathbb{N}} A_n$ de A de façon que

$$\sum_{n\in\mathbb{N}} |\mu(A_n)| > 2(1 + |\mu(A)|).$$

[1] On fait un abus de langage puisque dans la définition usuelle d'une partition les éléments de la partition sont tous non vides, ce qui n'est pas forcément le cas ici.

On a alors par exemple

$$\sum_{n\in\mathbb{N}} \mu(A_n)^+ > 1 + |\mu(A)|$$

(le cas symétrique $\sum_{n\in\mathbb{N}} \mu(A_n)^- > 1 + |\mu(A)|$ se traite de la même manière). On pose alors

$$B = \bigcup_{\{n:\mu(A_n)>0\}} A_n$$

de façon que

$$\mu(B) = \sum_{n\in\mathbb{N}} \mu(A_n)^+ > 1 + |\mu(A)|.$$

De plus, si $C = A\backslash B$,

$$|\mu(C)| = |\mu(A) - \mu(B)| \geq |\mu(B)| - |\mu(A)| > 1.$$

Par ailleurs, puisque $A = B \cup C$ et que $|\mu|$ est une mesure on doit avoir $|\mu|(B) = \infty$ ou $|\mu(C)| = \infty$, ce qui donne le résultat du lemme quitte à échanger les rôles de B et C si nécessaire. □

Nous pouvons maintenant compléter la preuve du théorème. On suppose que $|\mu|(E) = \infty$. Alors, on peut trouver des parties mesurables disjointes B_0 et C_0 avec $|\mu(B_0)| > 1$ et $|\mu|(C_0) = \infty$. En appliquant de même le lemme à C_0 on trouve B_1 et C_1 disjoints tels que $C_0 = B_1 \cup C_1$, $|\mu(B_1)| > 1$ et $|\mu|(C_1) = \infty$. Par récurrence, on construit ainsi une suite de parties mesurables disjointes $(B_n)_{n\in\mathbb{N}}$, telle que $|\mu(B_n)| > 1$ pour tout n. Cela contredit le fait que la série

$$\sum_{n\in\mathbb{N}} \mu(B_n)$$

doit converger absolument, d'après la définition d'une mesure signée. On conclut que $|\mu|(E) < \infty$. □

Exemple. Soit ν une mesure positive sur $(E, \mathcal{A})$, et soit $g \in L^1(E, \mathcal{A}, \nu)$. Alors la formule

$$\nu(A) = \int_A g\, d\nu$$

définit une mesure signée. En effet, si A est la réunion disjointe d'une suite (A_n) de parties mesurables, l'égalité

$$\mu(A) = \sum_{n\in\mathbb{N}} \mu(A_n)$$

est obtenue en observant que

$$g\, 1_A = \lim_{k\to\infty} g\, 1_{\cup_{n\leq k} A_n} \qquad \text{dans } L^1,$$

d'après le théorème de convergence dominée. Nous verrons plus loin que dans ce cas $|\mu| = |g| \cdot \nu$.

6.2 La décomposition de Jordan

Soit μ une mesure signée sur $(E, \mathcal{A})$. Alors, on vérifie immédiatement que les formules

$$\mu^+ = \frac{1}{2}(\mu + |\mu|),$$
$$\mu^- = \frac{1}{2}(|\mu| - \mu),$$

définissent deux mesures positives finies sur $(E, \mathcal{A})$. De plus, $\mu = \mu^+ - \mu^-$ et $|\mu| = \mu^+ + \mu^-$.

Théorème 6.2.1 *Soit μ une mesure signée sur $(E, \mathcal{A})$. Il existe une partie mesurable B de E, unique à un ensemble de $|\mu|$-mesure nulle près, telle que $\mu^+ = 1_B \cdot |\mu|$ et $\mu^- = 1_{B^c} \cdot |\mu|$ (de manière équivalente, μ^+, resp. μ^-, est la restriction de $|\mu|$ à B, resp. à B^c). De plus, on a pour tout $A \in \mathcal{A}$,*

$$\mu^+(A) = \mu^+(A \cap B) = \mu(A \cap B) \ , \quad \mu^-(A) = \mu^-(A \cap B^c) = -\mu(A \cap B^c).$$

En conséquence,

$$\mu(A) = \mu^+(A \cap B) - \mu^-(A \cap B^c),$$
$$|\mu|(A) = \mu^+(A \cap B) + \mu^-(A \cap B^c).$$

Preuve. On vérifie immédiatement que $\mu^+ \leq |\mu|$ et $\mu^- \leq |\mu|$, et donc les mesures μ^+ et μ^- sont absolument continues par rapport à $|\mu|$. D'après le théorème de Radon-Nikodym, il existe deux fonctions mesurables positives (finies) h_1 et h_2 telles que $\mu^+ = h_1 \cdot |\mu|$ et $\mu^- = h_2 \cdot |\mu|$. Puisque $\mu^+ \leq |\mu|$ et $\mu^- \leq |\mu|$, on sait que $0 \leq h_1 \leq 1$ et $0 \leq h_2 \leq 1$.

Si $h = h_1 - h_2$, on a alors, pour tout $A \in \mathcal{A}$,

$$\mu(A) = \mu^+(A) - \mu^-(A) = \int_A (h_1 - h_2)\, d|\mu|.$$

Il est facile de déduire de cette égalité que $|h_1 - h_2| = 1$, $|\mu|$ p.p. En effet, soit $r < 1$, et soit $(A_n)_{n\in\mathbb{N}}$ une partition mesurable de $E_r = \{x \in E : |h_1(x) - h_2(x)| \leq r\}$. Alors

$$\begin{aligned}
\sum_{n\in\mathbb{N}} |\mu(A_n)| &= \sum_{n\in\mathbb{N}} \Big| \int_{A_n} (h_1 - h_2) d|\mu| \Big| \\
&\leq \sum_{n\in\mathbb{N}} \int_{A_n} |h_1 - h_2| d|\mu| \\
&\leq \sum_{n\in\mathbb{N}} r\, |\mu|(A_n) \\
&= r\, |\mu|(E_r).
\end{aligned}$$

De la définition de $|\mu|$, il découle alors que $|\mu|(E_r) \leq r\, |\mu|(E_r)$, et donc $|\mu|(E_r) = 0$. Comme cela est vrai pour tout $r < 1$, on a $|h_1 - h_2| \geq 1$ μ p.p. et l'inégalité inverse est triviale.

Les propriétés $0 \leq h_1 \leq 1$, $0 \leq h_2 \leq 1$ et $|h_1 - h_2| = 1$ $|\mu|$ p.p. entraînent que

$$|\mu|(dx) \text{ p.p.} \qquad \begin{array}{l} \text{ou bien } h_1(x) = 1 \text{ et } h_2(x) = 0, \\ \text{ou bien } h_1(x) = 0 \text{ et } h_1(x) = 0. \end{array}$$

On pose alors $B = \{x \in E : h_1(x) = 1\}$. D'après ce qui précède on a $h_1 = 1_B$ et $h_2 = 1_{B^c}$, $|\mu|$ p.p. Cela donne les égalités $\mu^+ = 1_B \cdot |\mu|$ et $\mu^- = 1_{B^c} \cdot |\mu|$. L'unicité de B est une conséquence de l'unicité de la densité dans le théorème de Radon-Nikodym. Les autres propriétés de l'énoncé sont ensuite facilement établies. □

Remarque. Si $\mu = \mu_1 - \mu_2$ est une autre décomposition de μ comme différence de deux mesures positives finies, on a nécessairement $\mu_1 \geq \mu^+$ et $\mu_2 \geq \mu^-$. En effet,

$$\mu_1(A) \geq \mu_1(A \cap B) \geq \mu(A \cap B) = \mu^+(A \cap B) = \mu^+(A).$$

Intégration par rapport à une mesure signée.

Si $f \in \mathcal{L}^1(E, \mathcal{A}, |\mu|)$, on définit

$$\int f\, d\mu := \int f d\mu^+ - \int f d\mu^- = \int f(1_B - 1_{B^c}) d|\mu|.$$

Il est alors immédiat que

$$\Big| \int f\, d\mu \Big| \leq \int |f| d\mu.$$

Proposition 6.2.2 *Soit ν une mesure positive sur $(E, \mathcal{A})$, soit $g \in \mathcal{L}^1(E, \mathcal{A}, \nu)$, et soit μ la mesure signée définie par*

$$\mu(A) = \int_A g d\nu.$$

Alors $|\mu| = |g| \cdot \nu$. De plus, pour toute fonction $f \in \mathcal{L}^1(E, \mathcal{A}, |\mu|)$, on a $fg \in \mathcal{L}^1(E, \mathcal{A}, \nu))$, et

$$\int f\, d\mu = \int fg\, d\nu.$$

Preuve. Avec les notations du théorème précédent, on a pour tout $A \in \mathcal{A}$:

$$|\mu|(A) = \mu(A \cap B) - \mu(A \cap B^c) = \int_{A \cap B} g d\nu - \int_{A \cap B} g d\nu = \int_A gh\, d\nu,$$

en posant $h = 1_B - 1_{B^c}$. En prenant $A = \{x \in E : g(x)h(x) < 0\}$, on déduit facilement de cette égalité que $gh \geq 0$, ν p.p. Donc $gh = |gh| = |g|$, ν p.p., d'où

$$|\mu|(A) = \int_A |g| d\nu.$$

Ensuite, on a

$$\int |f| d|\mu| = \int |f|\, |g| d\nu$$

et donc $f \in \mathcal{L}^1(|\mu|) \Rightarrow fg \in \mathcal{L}^1(\nu)$. L'égalité

$$\int f d\mu = \int fg \, d\nu$$

est vraie par définition si f est étagée. Dans le cas, général, on utilise le fait qu'on peut écrire $f = \lim f_n$, où les fonctions f_n sont étagées et dominées en valeur absolue par $|f|$. Le théorème de convergence dominée appliqué à μ^+, μ^- et ν donne le résultat voulu. □

Le théorème de Radon-Nikodym pour les mesures signées.

Soit ν une mesure positive, et soit μ une mesure signée. On dit que μ est absolument continue par rapport à ν (notation : $\mu \ll \nu$) si

$$\forall A \in \mathcal{A}, \quad \nu(A) = 0 \Rightarrow \mu(A) = 0.$$

Théorème 6.2.3 *Soit μ une mesure signée et soit ν une mesure positive σ-finie. Les trois propriétés suivantes sont équivalentes :*

(i) $\mu \ll \nu$ *.*

(ii) *Pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que*

$$\forall A \in \mathcal{A}, \quad \nu(A) \leq \delta \Rightarrow |\mu|(A) \leq \varepsilon.$$

(iii) *Il existe $g \in \mathcal{L}^1(E, \mathcal{A}, \nu)$ telle que :*

$$\forall A \in \mathcal{A}, \quad \mu(A) = \int_A g d\nu.$$

Preuve. (ii)⇒(i) est évident. Montrons (i)⇒(iii). Si $\mu \ll \nu$, on aussi $\mu^+ \ll \nu$ et $\mu^- \ll \nu$, et donc le théorème de Radon-Nikodym pour les mesures positives permet d'écrire $\mu^+ = g_1 \cdot \nu$ et $\mu^- = g_2 \cdot \nu$ avec $g_1, g_2 \geq 0$, $\int g_1 d\nu = \mu^+(E) < \infty$ et $\int g_2 d\nu = \mu^-(E) < \infty$. On obtient ainsi (iii) avec $g = g_1 - g_2$.

Il reste à montrer (iii)⇒(ii). D'après la proposition précédente, on a $|\mu| = |g| \cdot \nu$. De plus, le théorème de convergence dominée entraîne que

$$\lim_{n\to\infty} \int_{\{|g|\geq n\}} |g| \, d\nu = 0.$$

Donc, si $\varepsilon > 0$ est fixé, on peut choisir N assez grand de façon que

$$\int_{\{|g|\geq N\}} |g| \, d\nu < \frac{\varepsilon}{2}.$$

Alors, en prenant $\delta = \varepsilon/(2N)$, on a, pour tout $A \in \mathcal{A}$ tel que $\nu(A) < \delta$,

$$|\mu|(A) = \int_A |g| d\nu \leq \int_{\{|g|\geq N\}} |g| \, d\nu + \int_{A\cap\{|g|<N\}} |g| \, d\nu \leq \frac{\varepsilon}{2} + N \, \frac{\varepsilon}{2N} = \varepsilon.$$

□

6.3 La dualité $L^p - L^q$

Soit ν une mesure positive sur $(E, \mathcal{A})$. Soit $p \in [1, \infty]$ et soit q l'exposant conjugué de p. Alors, si on fixe $g \in L^q(E, \mathcal{A}, \nu)$, la formule

$$\Phi_g(f) = \int f g \, d\nu$$

définit une forme linéaire continue sur $L^p(E, \mathcal{A}, \nu)$. En effet, l'inégalité de Hölder montre d'une part que $\Phi_g(f)$ est bien définie, d'autre part que

$$|\Phi_g(f)| \leq C_g \, \|f\|_p$$

avec $C_g = \|g\|_q$. On voit aussi que la norme opérateur de Φ_g, définie par

$$\|\Phi_g\| = \sup_{\|f\|_p \leq 1} |\Phi_g(f)|,$$

vérifie $\|\Phi_g\| \leq \|g\|_q$.

La question est alors de savoir si l'on obtient ainsi toutes les formes linéaires continues sur $L^p(E, \mathcal{A}, \nu)$ (dans le cas $p = q = 2$, la théorie des espaces de Hilbert nous dit déjà que la réponse est oui). Le théorème suivant donne la réponse lorsque $p < \infty$.

Théorème 6.3.1 *Soit ν une mesure σ-finie sur $(E, \mathcal{A})$, soit $p \in [1, \infty[$ et soit q l'exposant conjugué de p. Alors, si Φ est une forme linéaire continue sur $L^p(E, \mathcal{A}, \nu)$, il existe une unique $g \in L^q(E, \mathcal{A}, \nu)$ tel que, pour toute $f \in L^p(E, \mathcal{A}, \nu)$,*

$$\Phi(f) = \int f g \, d\nu.$$

De plus la norme opérateur de Φ est

$$\|\Phi\| = \|g\|_q.$$

Aves les notations précédant le théorème, on voit que l'application $g \longrightarrow \Phi_g$ permet d'identifier $L^q(\nu)$ au dual topologique de $L^p(\nu)$ (c'est à-dire à l'espace vectoriel des formes linéaires continues sur $L^p(\nu)$, muni de la norme opérateur). Nous verrons en remarque que cette propriété ne subsiste pas dans le cas $p = \infty$.

Preuve. Supposons d'abord $\nu(E) < \infty$. Alors, pour tout $A \in \mathcal{A}$, posons

$$\mu(A) = \Phi(1_A),$$

ce qui a bien un sens puisque $1_A \in L^p(\nu)$. On commence par vérifier que μ est une mesure signée sur $(E, \mathcal{A})$. Soit $(A_n)_{n \in \mathbb{N}}$ une famille dénombrable de parties mesurables disjointes. Si A désigne la réunion des A_n, on a

$$1_A = \lim_{k \to \infty} \sum_{n \leq k} 1_{A_n}$$

dans $L^p(\nu)$ (par convergence dominée, facilement justifiée puisque la fonction 1 est dans $L^p(\nu)$). En utilisant la continuité de Φ, on obtient ainsi

$$\mu(A) = \lim_{k\to\infty} \Phi\Big(\sum_{n\leq k} 1_{A_n}\Big) = \lim_{k\to\infty} \sum_{n\leq k} \mu(A_n).$$

La convergence absolue de la série $\sum \mu(A_n)$ est une conséquence : en notant $A'_n = A_n$ si $\mu(A_n) > 0$ et $A'_n = \varnothing$ sinon, et A' la réunion des A'_n, on a

$$\sum_n \mu(A_n)^+ = \sum_n \mu(A'_n) = \lim_{k\to\infty} \sum_{n\leq k} \mu(A'_n) = \mu(A') < \infty,$$

et de même pour les termes négatifs de la suite $(\mu(A_n))$. Une fois acquise la convergence absolue de la série, l'égalité $\mu(A) = \sum_n \mu(A_n)$ découle de ce qui précède.

Si $A \in \mathcal{A}$ et $\nu(A) = 0$, on a $1_A = 0$ dans $L^p(E, \mathcal{A}, \nu)$ et donc $\mu(A) = \Phi(1_A) = 0$. Donc $\mu \ll \nu$ et le théorème précédent montre qu'il existe une fonction $g \in L^1(E, \mathcal{A}, \nu)$ telle que

$$\forall A \in \mathcal{A}\ , \qquad \Phi(1_A) = \mu(A) = \int_A g\, d\nu.$$

L'égalité

$$\Phi(f) = \int f g\, d\nu$$

est vraie par linéarité lorsque f est étagée, puis lorsque f est seulement mesurable bornée puisqu'une telle fonction est limite uniforme (donc dans $L^p(\nu)$ parce que ν est finie) de fonctions étagées.

Montrons maintenant que $g \in L^q(\nu)$.

- Si $p = 1$, alors pour tout $A \in \mathcal{A}$,

$$\Big|\int_A g d\nu\Big| = |\Phi(1_A)| \leq \|\Phi\|\, \|1_A\|_1 = \|\Phi\|\, \nu(A)$$

  ce qui entraîne facilement que $|g| \leq \|\Phi\|$, ν p.p. (pour le voir considérer $A = \{g > \|\Phi\| + \varepsilon\}$ ou $A = \{g < -\|\Phi\| - \varepsilon\}$), et donc $\|g\|_\infty \leq \|\Phi\|$.

- Si $p \in]1, \infty[$, on pose $E_n = \{x \in E : |g(x)| \leq n\}$, puis $f_n = 1_{E_n} |g|^{q-1} \mathrm{signe}(g)$. Comme f_n est bornée, on a

$$\int_{E_n} |g|^q d\nu = \int f_n g\, d\nu = \Phi(f_n) \leq \|\Phi\|\, \|f_n\|_p = \|\Phi\| \Big(\int_{E_n} |g|^q d\nu\Big)^{1/p},$$

  d'où

$$\int_{E_n} |g|^q d\nu \leq \|\Phi\|^q.$$

  En faisant tendre n vers ∞, on trouve par convergence monotone que $\|g\|_q \leq \|\Phi\|$.

Dans les deux cas, on a obtenu que $g \in L^q(\nu)$ et $\|g\|_q \leq \|\Phi\|$. Vus comme fonctions de $f \in L^p(\nu)$, les deux membres de l'égalité

$$\Phi(f) = \int f g \, d\nu$$

sont des fonctions continues sur $L^p(\nu)$ qui coïncident lorsque f appartient au sous-ensemble dense des fonctions mesurables bornées. Elles coïncident donc partout.

Par ailleurs, comme expliqué avant l'énoncé de théorème, l'inégalité de Hölder entraîne que $\|\Phi\| \leq \|g\|_q$, et comme l'inégalité inverse a été obtenue ci-dessus, on a $\|\Phi\| = \|g\|_q$.

Enfin, l'application qui à $g \in L^q(\nu)$ associe la forme linéaire $f \longrightarrow \int f g \, d\nu$ est une isométrie de $L^q(\nu)$ sur le dual topologique de $L^p(\nu)$ (i.e. l'espace des formes linéaires continues sur $L^p(\nu)$) et est donc nécessairement injective. Cela donne l'unicité de g dans l'énoncé du théorème.

Il reste à traiter le cas $\nu(E) = \infty$. Dans ce cas, on peut écrire E comme la réunion d'une famille dénombrable disjointe $(E_n)_{n\in\mathbb{N}}$ de parties mesurables telles que $\nu(E_n) < \infty$ pour tout n. Notons ν_n la restriction de ν à E_n. Alors l'application $f \longrightarrow f 1_{E_n}$ induit une isométrie de $L^p(\nu_n)$ sur un sous-espace de $L^p(\nu)$. En remplaçant ν par ν_n on peut donc appliquer la première partie de la preuve à la forme linéaire continue Φ_n définie sur $L^p(\nu_n)$ par

$$\Phi_n(f) = \Phi(f 1_{E_n}).$$

Il existe donc une fonction $g_n \in L^q(\nu_n)$ telle que, pour toute fonction $f \in L^p(\nu_n)$,

$$\Phi(f 1_{E_n}) = \int f g_n \, d\nu_n.$$

Quitte à remplacer g_n par $g_n 1_{E_n}$ on peut supposer que $g_n = 0$ sur E_n^c, et réécrire le résultat précédent sous la forme

$$\Phi(f 1_{E_n}) = \int f g_n \, d\nu,$$

pour toute fonction $f \in L^p(\nu)$.

Si $f \in L^p(\nu)$, on a

$$f = \lim_{k\to\infty} \sum_{n\leq k} f \, 1_{E_n} \qquad \text{dans } L^p(\nu),$$

ce qui entraîne

$$\Phi(f) = \lim_{k\to\infty} \int f \Big(\sum_{n\leq k} g_n \Big) \, d\nu.$$

Par ailleurs, de l'inégalité

$$\int f \Big(\sum_{n\leq k} g_n \Big) d\nu = \Phi(f \sum_{n\leq k} 1_{E_n}) \leq \|\Phi\| \, \|f\|_p$$

on déduit grâce aux mêmes arguments que dans le cas où $\nu(E) < \infty$ que, pour tout entier $k \geq 1$,

$$\| \sum_{n\leq k} g_n \|_q \leq \|\Phi\|.$$

Posons maintenant pour tout $x \in E$,

$$g(x) = \sum_{n \in \mathbb{N}} g_n(x)$$

(il y a eu plus un terme non nul dans la somme pour chaque x). Si $q = \infty$, l'inégalité précédente montre que $\|g\|_\infty \leq \|\Phi\|$. Si $q < \infty$, la même inégalité donne

$$\int |g|^q d\nu = \sum_{n \in \mathbb{N}} \int |g_n|^q d\nu = \lim_{k \to \infty} \sum_{n \leq k} \int |g_n|^q d\nu \leq \|\Phi\|^q.$$

Dans les deux cas on a $g \in L^q(\nu)$. Enfin,

$$\Phi(f) = \lim_{k \to \infty} \int f\Big(\sum_{n \leq k} g_n\Big)\, d\nu = \int f g\, d\nu,$$

où dans la deuxième égalité l'application du théorème de convergence dominée est justifiée par la majoration $|\sum_{n \leq k} g_n| \leq |g|$.

L'égalité $\|\Phi\| = \|g\|_q$ et l'unicité de g sont maintenant obtenues par les mêmes arguments que dans le cas où $\nu(E) < \infty$. $\square$

Remarque. Lorsque $p = \infty$, le résultat du théorème est faux en général : il existe des formes linéaires continues sur $L^\infty(E, \mathcal{A}, \nu)$ qui ne peuvent pas se représenter sous la forme $\Phi(f) = \int f g\, d\nu$ avec une fonction $g \in L^1(E, \mathcal{A}, \nu)$. Considérons le cas de ℓ^∞, qui est l'espace des suites bornées $a = (a_k)_{k \in \mathbb{N}}$ de réels, muni de la norme $\|a\|_\infty = \sup a_k$. Soit H le sous-espace (fermé) de ℓ^∞ défini par

$$H = \{a \in \ell^\infty : \lim_{k \to \infty} a_k \text{ existe}\},$$

et définissons $\Phi : H \longrightarrow \mathbb{R}$ par

$$\Phi(a) = \lim_{k \to \infty} a_k.$$

Evidemment $|\Phi(a)| \leq \|a\|_\infty$. Le théorème de Hahn-Banach permet alors de prolonger Φ à une forme linéaire sur ℓ^∞, de façon que la propriété $|\Phi(a)| \leq \|a\|_\infty$ reste vraie pour tout $a \in \ell^\infty$. Il est facile de voir qu'on ne peut pas représenter Φ sous la forme

$$\Phi(a) = \sum_{k \in \mathbb{N}} a_k b_k$$

avec un élément $b = (b_k)_{k \in \mathbb{N}}$ de ℓ^1. En effet, si tel était le cas, en considérant pour tout $n \in \mathbb{N}$ l'élément $a^{(n)}$ de ℓ^∞ défini par $a_k^{(n)} = 1_{\{k=n\}}$, on trouverait, pour tout $n \in \mathbb{N}$,

$$b_n = \Phi(a^{(n)}) = 0,$$

ce qui est absurde.

6.4 Le théorème de représentation de Riesz

Dans tout ce paragraphe, nous supposons que E est un espace métrique localement compact séparable. On note $C_0(E)$ l'espace des fonctions continues sur E qui tendent vers 0 à l'infini : $f \in C_0(E)$ si et seulement si f est continue et si pour tout $\varepsilon > 0$ il existe un compact K de E tel que $|f(x)| < \varepsilon$ pour tout $x \in E \backslash K$. L'espace $C_0(E)$ est un espace de Banach pour la norme

$$\|f\| = \sup_{x \in E} |f(x)|.$$

Si μ est une mesure signée sur $(E, \mathcal{B}(E))$, l'application

$$\Phi(f) = \int_E f d\mu \ , \ f \in C_0(E),$$

définit une forme linéaire continue sur $C_0(E)$. De plus, cette forme linéaire est continue puisque

$$|\Phi(f)| \leq \int_E |f| \, d|\mu| \leq |\mu|(E) \, \|f\|.$$

Cette inégalité montre même que $\|\Phi\| \leq |\mu|(E)$.

Théorème 6.4.1 *Soit Φ une forme linéaire continue sur $C_0(E)$. Il existe alors une unique mesure signée μ sur $(E, \mathcal{B}(E))$ telle que*

$$\forall f \in C_0(E) \ , \qquad \Phi(f) = \int_E f \, d\mu.$$

Nous renvoyons au chapitre 6 de Rudin [7] pour une preuve qui traite en fait le cadre complexe plus général.

Remarque. L'espace $\mathcal{M}(E)$ des mesures signées sur E est un espace vectoriel, et il est facile de vérifier que l'application $\mu \longrightarrow |\mu|(E)$ définit une norme sur cet espace vectoriel. De plus, $\mathcal{M}(E)$ est complet pour cette norme. Le théorème précédent peut être alors reformulé en disant que $\mathcal{M}(E)$ est le dual topologique de $C_0(E)$.

Lorsque E est compact, l'espace $C_0(E)$ coïncide avec l'espace $C_b(E)$ des fonctions continues bornées sur E, et donc $\mathcal{M}(E)$ est le dual de $C_b(E)$. Cette assertion devient fausse lorsque E n'est pas compact, par exemple lorsque $E = \mathbb{R}$. Dans ce cas, il existe des formes linéaires continues sur $C_b(E)$ qui ne se représentent pas par des mesures signées (on peut en construire en adaptant l'exemple de la fin de la partie précédente).

Chapitre 7

Formule de changement de variables et compléments

La formule de changement de variables identifie l'image par un difféomorphisme de la mesure de Lebesgue sur un ouvert de $\mathbb{R}^d$. Après le théorème de Fubini, c'est le deuxième outil fondamental de calcul des intégrales. Comme application particulièrement importante, on donne ici la formule d'intégration en coordonnées polaires dans $\mathbb{R}^d$, ce qui conduit aussi à introduire la mesure de Lebesgue sur la sphère unité.

7.1 La formule de changement de variables

Nous commençons par traiter le cas particulier important d'une application affine.

Proposition 7.1.1 *Soit $b \in \mathbb{R}^d$ et soit M une matrice $d \times d$ à coefficients réels inversible. Définissons $f : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ par $f(x) = Mx + b$. Alors, pour tout borélien A de $\mathbb{R}^d$,*

$$\lambda(f(A)) = |\det(M)|\, \lambda(A).$$

Remarque. Si M n'est pas inversible, $f(A) \subset f(\mathbb{R}^d)$ est contenu dans un hyperplan, qui est de mesure de Lebesgue nulle (exercice !).

Preuve. Remarquons d'abord que $f(A) = (f^{-1})^{-1}(A) \in \mathcal{B}(\mathbb{R}^d)$ si $A \in \mathcal{B}(R^d)$. Grâce à l'invariance par translation de la mesure de Lebesgue, on se ramène au cas $b = 0$. Dans ce cas, on a pour tous $a \in \mathbb{R}^d$ et $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda(f(a + A)) = \lambda(f(a) + f(A)) = \lambda(f(A)),$$

ce qui montre que la mesure $A \longrightarrow \lambda(f(A))$ (mesure-image de λ par f^{-1}) est invariante par translation. Donc il existe une constante c telle que, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda(f(A)) = c\, \lambda(A).$$

Il reste à montrer que $c = |\det(M)|$.

Si M est une matrice orthogonale, et B_d désigne la boule unité fermée de $\mathbb{R}^d$, on a $f(B_d) = B_d$, d'où il découle aussitôt que $c = 1 = |\det(M)|$ dans ce cas.

Si M est une matrice symétrique définie positive, alors on peut trouver une matrice orthogonale P telle que ${}^tPMP$ soit diagonale avec coefficients diagonaux $\alpha_i > 0$, $i \in \{1, \ldots, d\}$. Alors,

$$f(P([0,1]^d)) = \{MPx : x \in [0,1]^d\} = \{Py : y \in \prod_{i=1}^{d} [0, \alpha_i]\},$$

et donc, en utilisant le cas orthogonal,

$$c = c\,\lambda(P([0,1]^d)) = \lambda(f(P([0,1]^d))) = \lambda\Big(\{Py : y \in \prod_{i=1}^{d} [0, \alpha_i]\}\Big) = \lambda\Big(\prod_{i=1}^{d} [0, \alpha_i]\Big) = \prod_{i=1}^{d} \alpha_i.$$

Dans ce cas on trouve encore $c = |\det(M)|$.

Enfin, dans le cas général, on remarque qu'on peut écrire $M = PS$, où P est orthogonale et S est symétrique définie positive (prendre $S = \sqrt{{}^tMM}$ et $P = MS^{-1}$). En utilisant les deux cas particuliers ci-dessus, on trouve aussitôt :

$$c = |\det(P)|\,|\det(S)| = |\det(M)|.$$

□

Soient U et D deux ouverts de $\mathbb{R}^d$. On dit qu'une application $\varphi : U \longrightarrow D$ est un difféomorphisme de classe C^1 si φ est bijective et de classe C^1 sur U et si φ^{-1} est aussi de classe C^1 sur D. On sait qu'alors la dérivée $\varphi'(u)$ est inversible, pour tout $u \in U$.

Théorème 7.1.2 *Soit $\varphi : U \longrightarrow D$ un difféomorphisme de classe C^1. Alors pour toute fonction borélienne $f : D \longrightarrow \mathbb{R}_+$,*

$$\int_D f(x)\,dx = \int_U f(\varphi(u))\,|J_\varphi(u)|\,du\,,$$

où $J_\varphi(u) = \det(\varphi'(u))$ est le Jacobien de φ en u.

Preuve. Par les arguments habituels (passage à la limite croissant) on se ramène au cas où f est étagée positive, puis au cas $f = 1_A$, A étant un borélien de D. Dans ce cas, l'égalité du théorème s'écrit :

$$\lambda(A) = \int_{\varphi^{-1}(A)} |J_\varphi(u)|\,du.$$

Quitte à remplacer A par $\varphi^{-1}(A)$, il suffit de montrer que, pour tout borélien A de U,

$$\lambda(\varphi(A)) = \int_A |J_\varphi(u)|\,du. \tag{7.1}$$

(Remarquer que $\varphi(A) = (\varphi^{-1})^{-1}(A)$ est borélien.)

Lemme 7.1.3 *Soit K un compact de U et soit $\varepsilon > 0$. Alors on peut choisir $\delta > 0$ assez petit de manière que, pour tout cube C de faces parallèles aux axes, de centre $u_0 \in K$ et de coté de longueur inférieure à δ,*

$$(1-\varepsilon)|J_\varphi(u_0)|\,\lambda(C) \leq \lambda(\varphi(C)) \leq (1+\varepsilon)|J_\varphi(u_0)|\,\lambda(C).$$

Preuve du lemme. En utilisant la continuité de φ', on voit qu'on peut choisir $\delta > 0$ assez petit pour que d'une part $\delta < \frac{1}{d}\operatorname{dist}(K, U^c)$ et d'autre part, pour tout $u_0 \in K$ et tout $u \in \mathbb{R}^d$ tel que $|u - u_0| < d\delta$,

$$|\varphi(u) - \varphi(u_0) - \varphi'(u_0)\cdot(u-u_0)| \leq \varepsilon|u-u_0|.$$

Notons $f(v) = \varphi(u_0) + \varphi'(u_0)\cdot v$ pour $v \in \mathbb{R}^d$. On voit que, si $|u-u_0| < d\delta$,

$$\varphi(u) = f(u-u_0) + h(u,u_0),$$

avec $|h(u,u_0)| \leq \varepsilon|u-u_0|$. En prenant $g(u,u_0) = \varphi'(u_0)^{-1}\cdot h(u,u_0)$, on trouve que

$$\varphi(u) = f(u-u_0+g(u,u_0)),$$

où $|g(u,u_0)| \leq a\varepsilon|u-u_0|$, avec $a := \sup\{\|\varphi'(v)^{-1}\|;\, v \in K\} < \infty$.

Soit maintenant C un cube centré en u_0 et de coté $r \leq \delta$. Il découle de ce qui précède que

$$\varphi(C) \subset f((1+da\varepsilon)\widetilde{C}),$$

où $\widetilde{C}$ est le cube translaté de C centré en 0. Grâce à la proposition ci-dessus, il vient alors

$$\lambda(\varphi(C)) \leq \lambda(f((1+da\varepsilon)\widetilde{C})) = |\det\varphi'(u_0)|\,\lambda((1+da\varepsilon)\widetilde{C}) = (1+da\varepsilon)^d|J_\varphi(u_0)|\,\lambda(C),$$

ce qui donne la majoration souhaitée. La preuve de la minoration est analogue : on montre que pour une constante c' bien choisie, on a

$$\varphi^{-1}(f((1-c'\varepsilon)\widetilde{C})) \subset C\,,$$

d'où

$$f((1-c'\varepsilon)\widetilde{C}) \subset \varphi(C)$$

et on conclut de la même manière. □

On revient à la preuve du théorème. Soit $n \geq 1$ un entier. On appelle cube élémentaire d'ordre n tout cube de la forme

$$C = \prod_{j=1}^{d} \,]k_j 2^{-n}, (k_j+1)2^{-n}]\,, \quad k_j \in \mathbb{Z}.$$

On note $\mathcal{C}_n$ l'ensemble des cubes élémentaires d'ordre n.

Soit C_0 un cube élémentaire d'ordre n_0 fixé, tel que $\bar{C}_0 \subset U$, et soit $\varepsilon > 0$. Fixons $n \geq n_0$ assez grand pour que d'une part la conclusion du lemme soit vraie pour $K = \bar{C}_0$ et $\delta = 2^{-n}$, et d'autre part, pour tous $u, v \in K$ tels que $|u-v| \leq d\delta$,

$$(1-\varepsilon)|J_\varphi(u)| \leq |J_\varphi(v)| \leq (1+\varepsilon)|J_\varphi(u)|. \tag{7.2}$$

Alors, en notant x_C le centre d'un cube C,

$$\begin{aligned}
\lambda(\varphi(C_0)) &= \sum_{\substack{C\in\mathcal{C}_n\\ C\subset C_0}} \lambda(\varphi(C)) \\
&\leq (1+\varepsilon) \sum_{\substack{C\in\mathcal{C}_n\\ C\subset C_0}} |J_\varphi(x_C)|\,\lambda(C) \\
&\leq (1+\varepsilon)^2 \sum_{\substack{C\in\mathcal{C}_n\\ C\subset C_0}} \int_C |J_\varphi(u)|\,du \\
&= (1+\varepsilon)^2 \int_{C_0} |J_\varphi(u)|\,du.
\end{aligned}$$

On a utilisé le lemme dans la première inégalité, et (7.2) dans la seconde. On obtient de même la minoration

$$\lambda(\varphi(C_0)) \geq (1-\varepsilon)^2 \int_{C_0} |J_\varphi(u)|\,du.$$

Comme ε était arbitraire, on conclut que

$$\lambda(\varphi(C_0)) = \int_{C_0} |J_\varphi(u)|\,du.$$

On a donc obtenu (7.1) lorsque A est un cube élémentaire d'adhérence contenue dans A. Le cas général découle maintenant d'arguments de classe monotone. Notons μ la mesure-image de la mesure de Lebesgue sur D par φ^{-1} :

$$\mu(A) = \lambda(\varphi(A))$$

pour tout borélien A de U. Soit aussi

$$\widetilde{\mu}(A) = \int_A |J_\varphi(u)|\,du.$$

On a obtenu que $\mu(C) = \widetilde{\mu}(C)$ pour tout cube élémentaire C d'adhérence contenue dans U. D'autre part, si U_n désigne la réunion (disjointe) des cubes élémentaires d'ordre n d'adhérence contenue dans $U \cap \{u : |u| \leq n\}$, on a $U_n \uparrow U$ quand $n \to \infty$ et $\mu(U_n) = \widetilde{\mu}(U_n) < \infty$ pour tout n. Comme la classe des cubes élémentaires d'adhérence contenue dans U est stable par intersection finie et engendre la tribu borélienne $\mathcal{B}(U)$, on peut appliquer le dernier corollaire du Chapitre 1 pour conclure que $\mu = \widetilde{\mu}$, ce qui était le résultat recherché. $\square$

Application à l'intégrale en coordonnées polaires.

On prend $d = 2$, $U =]0,\infty[\times]-\pi,\pi[$ et $D = \mathbb{R}^2 \backslash \{(x,0); x \leq 0\}$. Alors l'application

$$\varphi(r,\theta) = (r\cos\theta, r\sin\theta)\ , \quad (r,\theta) \in U$$

est un difféomorphisme de classe C^1 de U sur D. On calcule facilement

$$\varphi'(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

et donc $J_\varphi(r,\theta) = r$.

Il découle du théorème que, pour toute fonction borélienne $f : \mathbb{R}^2 \longrightarrow \mathbb{R}_+$,

$$\int_D f(x,y)\,dxdy = \int_U f(r\cos\theta, r\sin\theta)\,r\,drd\theta = \int_0^\infty \int_{-\pi}^{\pi} f(r\cos\theta, r\sin\theta)\,r\,drd\theta.$$

Comme la demi-droite négative est de mesure de Lebesgue nulle dans $\mathbb{R}^2$, on a aussi

$$\int_{\mathbb{R}^2} f(x,y)\,dxdy = \int_0^\infty \int_{-\pi}^{\pi} f(r\cos\theta, r\sin\theta)\,r\,drd\theta.$$

Exemple. Pour $f(x,y) = \exp(-x^2-y^2)$, le théorème de Fubini-Tonnelli donne d'une part

$$\int_{\mathbb{R}^2} e^{-x^2-y^2}\,dxdy = \Big(\int_{-\infty}^{+\infty} e^{-x^2}\,dx\Big)^2$$

et d'autre part

$$\int_0^\infty \int_{-\pi}^{\pi} f(r\cos\theta, r\sin\theta)\,r\,drd\theta = 2\pi \int_0^\infty e^{-r^2}\,r\,dr = \pi,$$

ce qui donne la valeur

$$\int_{-\infty}^{+\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

7.2 Mesure de Lebesgue sur la sphère unité

Dans cette partie on note λ_d la mesure de Lebesgue sur $\mathbb{R}^d$. Soit S^{d-1} la sphère unité de $\mathbb{R}^d$:

$$S^{d-1} = \{x \in \mathbb{R}^d : |x| = 1\}.$$

Si $A \in \mathcal{B}(S^{d-1})$, on note $\Gamma(A)$ le borélien de $\mathbb{R}^d$ défini par

$$\Gamma(A) = \{rx;\ r \in [0,1] \text{ et } x \in A\}.$$

Théorème 7.2.1 *Pour tout $A \in \mathcal{B}(S^{d-1})$, on pose*

$$\omega_d(A) = d\,\lambda_d(\Gamma(A)).$$

Alors ω_d est une mesure positive finie sur S^{d-1}, qui est invariante par les isométries vectorielles. De plus, pour toute fonction borélienne $f : \mathbb{R}^d \longrightarrow \mathbb{R}_+$,

$$\int_{\mathbb{R}^d} f(x)\,dx = \int_0^\infty \int_{S^{d-1}} f(rz)\,r^{d-1}\,dr\,\omega_d(dz). \tag{7.3}$$

Enfin la masse totale de ω_d (volume de la sphère unité) est

$$\omega_d(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)}.$$

Remarque. On peut aussi montrer que toute mesure finie sur S^{d-1} invariante par les isométries vectorielles est proportionnelle à ω_d.

Preuve. Il est immédiat que ω_d est une mesure positive finie sur S^{d-1} : on peut la voir comme l'image de la restriction de $d\,\lambda_d$ à la boule unité B_d par l'application $x \longrightarrow \frac{x}{|x|}$. Le fait que λ_d soit invariante par les isométries vectorielles de $\mathbb{R}^d$ (proposition de la partie 1) entraîne facilement que ω_d l'est aussi. En effet, si φ est une telle isométrie,

$$\lambda_d(\Gamma(\varphi^{-1}(A))) = \lambda_d(\varphi^{-1}(\Gamma(A))) = \lambda_d(\Gamma(A)).$$

La masse totale de ω_d est

$$\omega_d(S^{d-1}) = d\,\lambda_d(B_d) = d\,\frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})}.$$

Il reste à établir (7.3). Il suffit de traiter le cas $f = 1_B$, où B est un borélien de $\mathbb{R}^d\backslash\{0\}$. La formule

$$\mu(B) = \int_0^\infty \int_{S^{d-1}} 1_B(rz)\, r^{d-1}\, dr\, \omega_d(dz)$$

définit une mesure μ sur $\mathbb{R}^d\backslash\{0\}$ et le problème est de montrer que $\mu = \lambda_d$. Considérons d'abord le cas où B est de la forme

$$B = \{x \in \mathbb{R}^d\backslash\{0\};\ a < |x| \le b \text{ et } \frac{x}{|x|} \in A\},$$

où A est un borélien de S^{d-1}, et $0 < a \le b$. Alors,

$$\mu(B) = \omega_d(A) \int_a^b r^{d-1}\, dr = \frac{b^d - a^d}{d}\omega_d(A).$$

Pour calculer $\lambda_d(B)$, notons $\alpha = \frac{a}{b} \in]0,1[$, et pour tout entier $n \ge 0$ posons

$$\Gamma_n(A) = \{y = rx;\ \alpha^{n+1} < r \le \alpha^n \text{ et } x \in A\}.$$

Alors, $\lambda_d(\Gamma_n(A)) = \alpha^{nd}\lambda_d(\Gamma_0(A))$ et par ailleurs

$$\lambda_d(\Gamma(A)) = \sum_{n=0}^{\infty} \lambda_d(\Gamma_n(A)).$$

Il en découle aussitôt que

$$\lambda_d(\Gamma_0(A)) = (1-\alpha^d)\,\lambda_d(\Gamma(A)) = \frac{1-\alpha^d}{d}\,\omega_d(A),$$

et puisque $B = b\,\Gamma_0(A)$,

$$\lambda_d(B) = b^d \lambda_d(\Gamma_0(A)) = \frac{b^d - a^d}{d}\omega_d(A) = \mu(B).$$

Finalement, la classe des ensembles B de la forme ci-dessus est stable par intersections finies, et on voit facilement qu'elle engendre la tribu borélienne sur $\mathbb{R}^d\backslash\{0\}$. Les arguments de classe monotone habituels montrent alors que $\mu = \lambda_d$. $\square$

Si $f : \mathbb{R}^d \longrightarrow \mathbb{R}_+$ est une fonction radiale, au sens où $f(x) = f(|x|)$, le théorème montre que

$$\int_{\mathbb{R}^d} f(x)\,dx = c_d \int_0^\infty f(r)\,r^{d-1}\,dr,$$

avec $c_d = \omega_d(S^{d-1})$.

Partie II

Probabilités

Chapitre 8

Fondements de la théorie des probabilités

Ce chapitre introduit les notions fondamentales de la théorie des probabilités : variables aléatoires, espérance, loi, moments de variables aléatoires, fonctions caractéristiques, etc. Puisque un espace de probabilité n'est rien d'autre qu'un espace mesurable muni d'une mesure de masse totale 1, beaucoup de ces notions correspondent à ce qui a déjà été vu dans le cadre de la théorie de l'intégration. Par exemple une variable aléatoire n'est rien d'autre qu'une fonction mesurable, et la notion d'espérance coïncide avec l'intégrale. Cependant, le point de vue de la théorie des probabilités, qui est expliqué ci-dessous, est bien différent, et une difficulté importante est de comprendre ce point de vue. Ainsi, la notion de loi, qui est un cas particulier de la notion de mesure-image, devient-elle maintenant fondamentale car elle permet d'évaluer la probabilité qu'une variable aléatoire "tombe" dans un ensemble donné.

8.1 Définitions générales

8.1.1 Espaces de probabilité

Soit $(\Omega, \mathcal{A})$ un espace mesurable, et soit P une mesure de probabilité sur $(\Omega, \mathcal{A})$. On dit alors que $(\Omega, \mathcal{A})$ est un espace de probabilité.

Un espace de probabilité est donc un cas particulier d'espace mesuré, pour lequel la masse totale de la mesure est égale à 1. En fait, le point de vue diffère de la théorie de l'intégration : dans le cadre de la théorie des probabilités, on cherche à fournir un modèle mathématique pour une "expérience aléatoire".

- Ω représente l'ensemble de toutes les éventualités possibles, toutes les déterminations du hasard dans l'expérience considérée.

- $\mathcal{A}$ est l'ensemble des "événements", qui sont les parties de Ω dont on peut évaluer la probabilité. Il faut voir un événement $A \in \mathcal{A}$ comme un sous-ensemble de Ω contenant toutes les éventualités ω pour lesquelles une certaine propriété est vérifiée.

- Pour $A \in \mathcal{A}$, $P(A)$ représente la probabilité d'occurrence de l'événement A. Dans les premiers traités de théorie des probabilités, longtemps avant l'introduction de la théorie de la mesure, la probabilité $P(A)$ était définie de la manière suivante : on imagine qu'on répète l'expérience aléatoire un nombre N de fois, et on note N_A le nombre de répétitions pour lesquelles l'événement A est réalisé; alors, la proportion N_A/N converge quand $N \to \infty$ vers la probabilité $P(A)$. Nous verrons plus loin le lien entre cette définition "historique" et l'approche moderne.

Exemples. (1) On lance un dé deux fois :

$$\Omega = \{1, 2, \ldots, 6\}^2 \ , \ \mathcal{A} = \mathcal{P}(\Omega) \ , \ P(A) = \frac{\text{Card}(A)}{36}.$$

Le choix de la probabilité correspond à l'idée que tous les résultats possibles pour les deux tirages sont équiprobables.

(2) On lance le dé jusqu'à obtenir un 6. Ici le choix de Ω est déjà moins évident. Comme le nombre de lancers nécessaires n'est a priori pas borné, le bon choix est d'imaginer qu'on fait une infinité de lancers :

$$\Omega = \{1, 2, \ldots, 6\}^{\mathbb{N}^*}$$

de sorte qu'un élément de Ω est une suite $\omega = (\omega_1, \omega_2, \ldots)$ qui donne les résultats des tirages successifs. La tribu $\mathcal{A}$ sur Ω est la tribu-produit définie comme la plus petite tribu rendant mesurables tous les ensembles de la forme

$$\{\omega : \omega_1 = i_1, \omega_2 = i_2, \ldots, \omega_n = i_n\}$$

où $n \geq 1$ et $i_1, \ldots, i_n \in \{1, 2, \ldots, 6\}$ ($\mathcal{A}$ coïncide aussi avec la tribu borélienne pour la topologie produit sur Ω). Enfin P est l'unique mesure de probabilité sur Ω telle que, pour tout choix de n et de $i_1, \ldots, i_n$,

$$P(\{\omega : \omega_1 = i_1, \omega_2 = i_2, \ldots, \omega_n = i_n\}) = (\frac{1}{6})^n.$$

L'unicité de P est une conséquence simple du lemme de classe monotone. L'existence est un cas particulier de la construction de mesures sur des produits infinis. On peut aussi construire P facilement partir de la mesure de Lebesgue sur $[0, 1]$: si à tout réel $x \in [0, 1]$ on associe la suite $(\varepsilon_k)_{k \in \mathbb{N}^*} \in \Omega$ telle que $x = \sum_{k=1}^{\infty} (\varepsilon_k - 1)\, 6^{-k}$ (cette suite est unique pour presque tout x), la probabilité P est obtenue comme mesure-image de la mesure de Lebesgue sur $[0, 1]$ par l'application $x \longrightarrow (\varepsilon_k)_{k \in \mathbb{N}^*}$.

(3) On s'intéresse au déplacement dans l'espace d'une particule ponctuelle soumise à des perturbations aléatoires. Si on se limite à l'intervalle de temps $[0, 1]$, l'espace de probabilité naturel est $C([0, 1], \mathbb{R}^3)$: un élément de Ω, une trajectoire possible, est une fonction continue $\omega : [0, 1] \longrightarrow \mathbb{R}^3$. La tribu sur Ω est alors la plus petite tribu qui rende mesurables toutes les applications coordonnées $\omega \longrightarrow \omega(t)$ pour $t \in \mathbb{R}_+$. Cette tribu coïncide avec la tribu borélienne pour la topologie de la convergence uniforme sur Ω. Il resterait à construire la probabilité P, pour laquelle de multiples choix sont possibles. L'exemple le plus important,

à la fois du point de vue théorique et pour les applications, est la mesure de Wiener, qui est la loi du mouvement brownien.

Remarque importante. Très souvent dans la suite, on ne spécifiera pas le choix de l'espace de probabilité. Les données importantes seront les propriétés des fonctions définies sur cet espace, les variables aléatoires.

8.1.2 Variables aléatoires

Définition 8.1.1 *Soit $(E, \mathcal{E})$ un espace mesurable. Une application mesurable $X : \Omega \longrightarrow E$ est appelée variable aléatoire (v.a. en abrégé) à valeurs dans E.*

Exemples. En reprenant les trois exemples ci-dessus :

(1) $X((i,j)) = i + j$ définit une variable aléatoire à valeurs dans $\{1, 2, \ldots, 12\}$.

(2) $X(\omega) = \inf\{j : \omega_j = 6\}$, avec la convention $\inf \varnothing = \infty$, définit une v.a. à valeurs dans $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. Pour vérifier la mesurabilité, on observe que, pour tout $k \geq 1$,

$$X^{-1}(\{k\}) = \{\omega \in \Omega : \omega_1 \neq 6, \omega_2 \neq 6, \ldots, \omega_{k-1} \neq 6, \omega_k = 6\}.$$

(3) Pour $t \in [0,1]$ fixé, $X(\omega) = \omega(t)$ est une v.a. à valeurs dans $\mathbb{R}^3$. (Remarquons que nous n'avons pas construit P dans cet exemple, mais cela n'intervient pas pour les questions de mesurabilité.)

Définition 8.1.2 *La loi de la variable aléatoire X est la mesure-image de P par X. C'est donc la mesure de probabilité sur $(E, \mathcal{E})$, notée P_X, définie par*

$$P_X(B) = P(X^{-1}(B)) \ , \qquad \forall B \in \mathcal{E}.$$

En pratique on écrit plutôt :

$$P_X(B) = P(X \in B) \ (= P(\{\omega \in \Omega : X(\omega) \in B\}) \ ).$$

La loi P_X permet de calculer la probabilité des événements qui “dépendent” de la v.a. X. Il faut comprendre qu'à chaque $\omega \in \Omega$ on a associé un “point aléatoire” $X(\omega)$ dans E, et que $P_X(B)$ est la probabilité que ce point aléatoire tombe dans B.

Remarque. Si μ est une mesure de probabilité sur $\mathbb{R}^d$, ou sur un espace plus général, il existe une manière canonique de construire une variable aléatoire dont la loi est μ. Il suffit de prendre $\Omega = \mathbb{R}^d$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^d)$, $P = \mu$, puis de poser $X(\omega) = \omega$. La loi de X est μ, de manière évidente.

Cas particuliers.

• **Variables aléatoires discrètes**. C'est le cas où E est dénombrable (et $\mathcal{E}$ est l'ensemble des parties de E). La loi de X est alors

$$P_X = \sum_{x \in E} p_x \, \delta_x$$

où $p_x = P(X = x)$ et δ_x désigne la la mesure de Dirac en x. En effet,

$$P_X(B) = P(X \in B) = P(\Big(\bigcup_{x \in B} \{X = x\}\Big) = \sum_{x \in B} P(X = x) = \sum_{x \in E} p_x \, \delta_x(B).$$

En pratique, trouver la loi d'une v.a. discrète, c'est donc calculer toutes les probabilités $P(X = x)$ pour $x \in E$.

Exemple. Revenons à l'exemple (2) ci-dessus, avec $X(\omega) = \inf\{j : \omega_j = 6\}$. Alors, pour tout $k \geq 1$,

$$P(X = k) = P\Big(\bigcup_{i_1,\ldots,i_{k-1} \neq 6} \{\omega_1 = i_1, \ldots, \omega_{k-1} = i_{k-1}, \omega_k = 6\}\Big) = 5^{k-1} \, (\frac{1}{6})^k = \frac{1}{6}(\frac{5}{6})^{k-1}.$$

Remarquons que $\sum_{k=1}^{\infty} P(X = k) = 1$ et donc $P(X = \infty) = 1 - P(X \in \mathbb{N}) = 0$. Observons que l'ensemble $\{X = \infty\}$ est loin d'être vide puisqu'il contient toutes les suites $(i_1, i_2, \ldots)$ qui ne prennent pas la valeur 6.

• **Variables aléatoires à densité**. Une variable aléatoire X à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est dite à densité si P_X est absolument continue par rapport à la mesure de Lebesgue λ.

Dans ce cas, le théorème de Radon-Nikodym montre qu'il existe une fonction borélienne $p : \mathbb{R}^d \longrightarrow \mathbb{R}_+$ telle que

$$P_X(B) = \int_B p(x) \, dx.$$

On a en particulier $\int_{\mathbb{R}^d} p(x) dx = P(X \in \mathbb{R}^d) = 1$. La fonction p, qui est unique à en ensemble de mesure de Lebesgue nulle près, est appelée la densité de (la loi de) X.

Si $d = 1$, on a en particulier, pour tous $\alpha \leq \beta$,

$$P(\alpha \leq X \leq \beta) = \int_\alpha^\beta p(x) \, dx.$$

8.1.3 Espérance mathématique

Définition 8.1.3 *Soit X une variable aléatoire réelle (i.e. à valeurs dans $\mathbb{R}$). On note alors*

$$E[X] = \int_\Omega X(\omega) \, P(d\omega),$$

qui est bien définie dans les deux cas suivants :

· si $X \geq 0$ (alors $E[X] \in [0, \infty]$),

· si X est de signe quelconque et $E[|X|] = \int |X| dP < \infty$.

On étend cette définition au cas où $X = (X_1, \ldots, X_d)$ est une variable aléatoire à valeurs dans $\mathbb{R}^d$ en prenant alors $E[X] = (E[X_1], \ldots, E[X_d])$, pourvu bien sûr que chacune des espérances $E[X_i]$ soit bien définie.

Remarque. Si $X = 1_B$, $E[X] = P(B)$. En général, $E[X]$ s'interprète comme la moyenne de la v.a. X. Dans le cas particulier où Ω est fini et P attribue la même valeur à chaque singleton, $E[X]$ est bien la moyenne au sens usuel des valeurs prises par X.

Proposition 8.1.1 *Soit X une variable aléatoire à valeurs dans $(E, \mathcal{E})$. Pour toute fonction mesurable $f : E \longrightarrow [0, \infty]$, on a*

$$E[f(X)] = \int_E f(x)\, P_X(dx).$$

Preuve. C'est évidemment une propriété générale des mesures-images déjà rencontrée dans le cours d'intégration. On remarque que le résultat est vrai par définition pour $f = 1_B$ puis par linéarité pour toute fonction étagée positive. Dans le cas général, on utilise le théorème de convergence monotone et le fait que toute fonction mesurable positive est limite croissante d'une suite de fonctions étagées positives. □

Si f est de signe quelconque, la formule de la proposition reste vraie à condition que les intégrales soient bien définies, ce qui revient à $E[|f(X)|] < \infty$.

La donnée de P_X permet donc de calculer la valeur moyenne de variables aléatoires de la forme $f(X)$. Inversement, on utilise souvent la proposition pour calculer la loi d'une v.a. X : si on arrive à écrire

$$E[f(X)] = \int f\, d\nu$$

pour toute fonction f "suffisamment" générale, alors on peut identifier ν à la loi de X. Donnons un exemple simple de ce principe.

Proposition 8.1.2 *Soit $X = (X_1, \ldots, X_d)$ une v.a. à valeurs dans $\mathbb{R}^d$. Supposons que la loi de X a une densité $p(x_1, \ldots, x_d)$. Alors, pour tout $j \in \{1, \ldots, d\}$, la loi de X_j a une densité donnée par*

$$p_j(x) = \int_{\mathbb{R}^{d-1}} p(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_d)\, dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_d$$

(par exemple, si $d = 2$,

$$p_1(x) = \int_{\mathbb{R}} p(x, y)\, dy \;, \quad p_2(y) = \int_{\mathbb{R}} p(x, y)\, dx).$$

Preuve. Soit π_j la projection $\pi_j(x_1, \ldots, x_d) = x_j$. En utilisant le théorème de Fubini, on écrit, pour toute fonction borélienne $f : \mathbb{R} \longrightarrow \mathbb{R}_+$,

$$\begin{aligned} E[f(X_j)] = E[f(\pi_j(X))] &= \int_{\mathbb{R}^d} f(x_j) p(x_1, \ldots, x_d)\, dx_1 \ldots dx_d \\ &= \int_{\mathbb{R}^d} f(x_j) \Big(\int_{\mathbb{R}^{d-1}} p(x_1, \ldots, x_d)\, dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_d \Big) dx_j \\ &= \int_{\mathbb{R}} f(x_j) p_j(x_j)\, dx_j, \end{aligned}$$

ce qui donne le résultat voulu. □

Remarque. Si $X = (X_1, \dots, X_d)$ est une v.a. à valeurs dans $\mathbb{R}^d$, les lois P_{X_j}, qu'on appelle souvent les lois marginales de X, sont déterminées par la loi de X, simplement parce que $P_{X_j} = \pi_j(P_X)$, avec la notation ci-dessous. Il est important d'observer que :

la réciproque est fausse !

Pour un exemple, considérons une densité de probabilité q sur $\mathbb{R}$, et observons que la fonction $p(x_1, x_2) = q(x_1)q(x_2)$ est alors aussi une densité de probabilité sur $\mathbb{R}^2$. D'après une remarque ci-dessus on peut construire une v.a. $X = (X_1, X_2)$ à valeurs dans $\mathbb{R}^2$ dont la loi est la mesure de densité p par rapport à la mesure de Lebesgue. Mais alors les deux v.a. X et $X' = (X_1, X_1)$ ont mêmes lois marginales (la proposition ci-dessus montre que $P_{X_1}(dx) = P_{X_2}(dx) = q(x)dx$) alors que les lois P_X et $P_{X'}$ sont très différentes, simplement parce que $P_{X'}$ est portée par la diagonale de $\mathbb{R}^2$, qui est de mesure de Lebesgue nulle.

8.1.4 Exemple : le paradoxe de Bertrand

Pour illustrer les notions introduites dans les paragraphes précédents, considérons le problème suivant. On s'intéresse à la probabilité qu'une corde choisie au hasard sur un cercle ait une longueur plus grande que le coté du triangle équilatéral inscrit. Sans perte de généralité on peut supposer que le cercle est le cercle unité. Bertrand proposait deux méthodes de calcul :

(a) On choisit les deux extrémités de la corde au hasard sur le cercle. La première étant choisie, la longueur de la corde sera plus grande que le coté du triangle équilatéral inscrit si et seulement si la seconde extrémité est dans un secteur angulaire d'ouverture $2\pi/3$. La probabilité est donc $\frac{2\pi/3}{2\pi} = \frac{1}{3}$.

(b) On choisit le centre de la corde au hasard sur le disque unité. La probabilité désirée est la probabilité que le centre tombe dans le disque de rayon $1/2$ centré à l'origine. Comme l'aire de ce disque est un quart de l'aire du disque unité, on trouve comme probabilité $\frac{1}{4}$.

On obtient donc un résultat différent dans les deux cas. L'explication tient dans le fait que les deux méthodes correspondent à des expériences aléatoires différentes, représentées par des choix différents de l'espace de probabilité. Il n'y a donc aucune raison pour que la loi de la variable aléatoire que l'on considère (la longueur de la corde) soit la même dans les deux cas. Pour nous en convaincre, explicitons les choix des espaces de probabilité.

(a) Dans ce cas,

$$\Omega = [0, 2\pi[^2 \ , \quad \mathcal{A} = \mathcal{B}([0, 2\pi[^2) \ , \quad P(d\omega) = \frac{1}{4\pi^2} d\theta \, d\theta',$$

où on note $\omega = (\theta, \theta')$ pour $\omega \in \Omega$. La longueur de la corde est

$$X(\omega) = 2|\sin(\frac{\theta - \theta'}{2})|.$$

On calcule facilement la loi de X :

$$\begin{aligned} E[f(X)] &= \int_\Omega f(X(\omega))\, P(d\omega) \\ &= \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} f(2|\sin(\frac{\theta-\theta'}{2})|)\, d\theta d\theta' \\ &= \frac{1}{\pi}\int_0^{\pi} f(2\sin(\frac{u}{2}))\, du \\ &= \frac{1}{\pi}\int_0^2 f(x)\, \frac{1}{\sqrt{1-\frac{x^2}{4}}}\, dx. \end{aligned}$$

Donc X est une v.a. réelle à densité : $P_X(dx) = p(x)dx$, avec

$$p(x) = \frac{1}{\pi}\, \frac{1}{\sqrt{1-\frac{x^2}{4}}}\, 1_{[0,2]}(x).$$

En particulier, la probabilité recherchée est

$$P(X \geq \sqrt{3}) = \int_{\sqrt{3}}^2 p(x)\, dx = \frac{1}{3}.$$

(b) Maintenant,

$$\Omega = \{\omega = (y,z) \in \mathbb{R}^2 : y^2+z^2 < 1\}\ , \quad \mathcal{A} = \mathcal{B}(\Omega)\ , \quad P(d\omega) = \frac{1}{\pi}\, 1_\Omega(y,z)\, dy\, dz.$$

La longueur de la corde est

$$X(\omega) = 2\sqrt{1-y^2-z^2}$$

et pour calculer sa loi on écrit

$$\begin{aligned} E[f(X)] &= \frac{1}{\pi}\int_{\mathbb{R}^2} f(2\sqrt{1-y^2-z^2})\, 1_{\{y^2+z^2<1\}}\, dydz \\ &= 2\int_0^1 f(2\sqrt{1-r^2})\, r\, dr \\ &= \frac{1}{2}\int_0^2 f(x)\, x\, dx. \end{aligned}$$

Donc $P_X(dx) = p(x)dx$, avec

$$p(x) = \frac{1}{2}\, 1_{[0,2]}(x)\, x\, dx.$$

On peut remarquer que la densité obtenue est très différente de celle du cas (a). En particulier,

$$P(X \geq \sqrt{3}) = \int_{\sqrt{3}}^2 p(x)\, dx = \frac{1}{4}.$$

Exercice. Traiter le cas de la troisième méthode proposée par Bertrand : on choisit au hasard la direction du rayon orthogonal à la corde, puis le centre de la corde uniformément sur ce rayon.

8.1.5 Lois classiques

On donne dans ce paragraphe quelques exemples importants de lois.

Lois discrètes.

(a) *Loi uniforme.* Si E est un ensemble fini, $\mathrm{Card}(E) = n$, une v.a. X est de loi uniforme sur E si

$$P(X = x) = \frac{1}{n}\ , \quad \forall x \in E.$$

(b) *Loi de Bernoulli de paramètre* $p \in [0,1]$. C'est la loi d'une v.a. X à valeurs dans $\{0,1\}$ telle que

$$P(X = 1) = p\ ,\ P(X = 0) = 1 - p.$$

On interprète X comme le résultat du lancer d'une pièce truquée qui tombe sur pile avec probabilité p.

(c) *Loi binômiale* $\mathcal{B}(n,p)$ ($n \in \mathbb{N}^*$, $p \in [0,1]$). C'est la loi d'une v.a. X à valeurs dans $\{1,\ldots,n\}$ telle que

$$P(X = k) = C_n^k\, p^k\, (1-p)^{n-k}.$$

On interprète X comme le nombre de piles obtenus en n lancers avec la pièce précédente.

(d) *Loi géométrique de paramètre* $p \in]0,1[$. C'est la loi d'une v.a. X à valeurs dans $\mathbb{N}$, telle que

$$P(X = k) = (1-p)\, p^k.$$

X est le nombre de piles obtenus avant le premier face.

(e) *Loi de Poisson de paramètre* $\lambda > 0$. X est une v.a. à valeurs dans $\mathbb{N}$, et

$$P(X = k) = \frac{\lambda^k}{k!}\, e^{-\lambda}\ , \quad \forall k \in \mathbb{N}.$$

On calcule facilement $E[X] = \lambda$. La loi de Poisson est très importante aussi bien du point de vue théorique que dans les applications. Intuitivement, elle correspond au nombre d'événements rares qui se sont produits durant une période longue. La traduction mathématique de cette intuition est l'*approximation binômiale de la loi de Poisson* : si pour tout $n \geq 1$, X_n suit une loi binômiale $\mathcal{B}(n, p_n)$ et si $np_n \longrightarrow \lambda$ quand $n \to \infty$, alors pour tout entier $k \in \mathbb{N}$,

$$\lim_{n\to\infty} P(X_n = k) = \frac{\lambda^k}{k!}\, e^{-\lambda}.$$

Lois continues. Dans les trois exemples qui suivent, X est une v.a. à valeurs dans $\mathbb{R}$, à densité $p(x)$.

(a) *Loi uniforme sur* $[a,b]$ $(a < b)$.

$$p(x) = \frac{1}{b-a} 1_{[a,b]}(x).$$

(b) *Loi exponentielle de paramètre* $\lambda > 0$.

$$p(x) = \lambda\, e^{-\lambda x}\, 1_{\mathbb{R}_+}(x).$$

Les lois exponentielles possèdent la propriété caractéristique suivante : si $a, b > 0$,

$$P(X > a + b) = P(X > a)\, P(X > b),$$

ce qu'on interprète en disant que la probabilité que $X - a > b$ sachant que $X > a$ coïncide avec la probabilité que $X > b$. C'est la propriété d'absence de mémoire de la loi exponentielle, qui explique qu'elle soit utilisée par exemple pour modéliser les temps de vie de machine sans usure.

(c) *Loi gaussienne, ou normale,* $\mathcal{N}(m, \sigma^2)$ ($m \in \mathbb{R}$, $\sigma > 0$).

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(- \frac{(x-m)^2}{2\sigma^2}\Big).$$

Avec la loi de Poisson, c'est la loi la plus importante en théorie des probabilités. Sa densité est la fameuse courbe en cloche. Les paramètres m et σ s'interprètent comme

$$m = E[X]\ , \quad \sigma^2 = E[(X-m)^2].$$

On remarque aussi que $X - m$ suit la loi $\mathcal{N}(0, \sigma^2)$. La loi gaussienne jouera un rôle important dans le Chapitre 10.

Par convention on dira qu'une v.a. constante égale à m suit la loi gaussienne $\mathcal{N}(m, 0)$. Si X suit la loi $\mathcal{N}(m, \sigma^2)$, pour tous $\lambda, \mu \in \mathbb{R}$, $\lambda X + \mu$ suit la loi $\mathcal{N}(\lambda m + \mu, \lambda^2\sigma^2)$.

8.1.6 Fonction de répartition d'une variable aléatoire réelle

Si X est une v.a. réelle, la *fonction de répartition* de X est la fonction $F_X : \mathbb{R} \longrightarrow [0, 1]$ définie par

$$F_X(t) = P(X \leq t) = P_X(]-\infty, t])\ , \quad \forall t \in \mathbb{R}.$$

La fonction F_X est croissante, continue à droite et a pour limite 0 en $-\infty$ et 1 en $+\infty$.

Inversement, si on se donne une fonction F ayant ces propriétés, on a vu dans le cours d'intégration qu'il existe une (unique) mesure de probabilité μ telle que $\mu(]-\infty, t]) = F(t)$ pour tout $t \in \mathbb{R}$. Cela montre qu'on peut interpréter F comme la fonction de répartition d'une v.a. réelle.

Il découle des résultats du cours d'intégration que F_X caractérise la loi P_X de X. On a en particulier

$$\begin{aligned} P(a \leq X \leq b) &= F_X(b) - F_X(a-) \qquad \text{si } a \leq b,\\ P(a < X < b) &= F_X(b-) - F_X(a) \qquad \text{si } a < b, \end{aligned}$$

et les sauts de F_X correspondent aux atomes de P_X.

8.1.7 Tribu engendrée par une variable aléatoire

Soit X une v.a. à valeurs dans un espace mesurable quelconque $(E,\mathcal{E})$. La *tribu engendrée par* X, notée $\sigma(X)$, est par définition la plus petite tribu sur Ω qui rende X mesurable :

$$\sigma(X) = \{A = X^{-1}(B) : B \in \mathcal{E}\}.$$

Remarque. On peut généraliser cette définition à une famille quelconque $(X_i)_{i\in I}$ de v.a., X_i étant à valeurs dans $(E_i,\mathcal{E}_i)$. Dans ce cas,

$$\sigma(X) = \sigma(X_i^{-1}(B_i) : B_i \in \mathcal{E}_i, i \in I).$$

Proposition 8.1.3 *Soit X une variable aléatoire à valeurs dans $(E,\mathcal{E})$, et soit Y une v.a. réelle. Il y a équivalence entre :*

(i) Y est $\sigma(X)$-mesurable.

(ii) Il existe une fonction mesurable f de $(E,\mathcal{E})$ dans $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ telle que $Y = f(X)$.

Preuve. L'implication (ii)⇒(i) est facile puisqu'une composée de fonctions mesurables est mesurable.

Dans l'autre sens, supposons que Y est $\sigma(X)$-mesurable. Traitons d'abord le cas où Y est étagée :

$$Y = \sum_{i=1}^{n} \lambda_i \, 1_{A_i}$$

où $\lambda_i \in \mathbb{R}$ et $A_i \in \sigma(X)$, pour tout $i \in \{1,\dots,n\}$. Alors, pour chaque $i \in \{1,\dots,n\}$, on peut trouver $B_i \in \mathcal{E}$ tel que $A_i = X^{-1}(B_i)$, et on a

$$Y = \sum_{i=1}^{n} \lambda_i \, 1_{A_i} = \sum_{i=1}^{n} \lambda_i \, 1_{B_i} \circ X = f \circ X,$$

où $f = \sum_{i=1}^{n} \lambda_i \, 1_{B_i}$ est $\mathcal{E}$-mesurable.

Dans le cas général, on sait que Y est limite simple d'une suite de v.a. Y_n étagées et $\sigma(X)$-mesurables. D'après la première étape, on peut écrire, pour tout n, $Y_n = f_n(X)$, où la fonction $f_n : E \longrightarrow \mathbb{R}$ est mesurable. On pose alors pour tout $x \in E$:

$$f(x) = \begin{cases} \lim\limits_{n\to\infty} f_n(x) & \text{si la limite existe,} \\ 0 & \text{sinon.} \end{cases}$$

On sait que la fonction f ainsi définie est mesurable. Par ailleurs, pour tout $\omega \in \Omega$, $X(\omega)$ appartient à l'ensemble des x pour lesquels $\lim f_n(x)$ existe (puisque $\lim f_n(X(\omega)) = \lim Y_n(\omega) = Y(\omega)$), et de plus

$$f(X(\omega)) = \lim f_n(X(\omega)) = Y(\omega)$$

ce qui donne la représentation recherchée $Y = f(X)$. □

8.2 Moments de variables aléatoires

8.2.1 Moments d'ordre p et variance

Soit X une v.a. réelle et soit $p \geq 1$ un entier. Le *moment d'ordre* p de X est par définition la quantité $E[X^p]$, qui n'est définie que si $E[|X|^p] < \infty$, ou si $X \geq 0$. La quantité $E[|X|^p]$ est appelée moment absolu d'ordre p. En particulier le moment d'ordre 1 est simplement l'espérance de X. On dit que la v.a. réelle X est *centrée* si elle est intégrable et si $E[X] = 0$.

L'espérance mathématique est un cas particulier d'intégrale par rapport à une mesure positive, et on peut donc lui appliquer les théorèmes généraux vus dans ce cadre. En particulier, les théorèmes de convergence sont d'un usage fréquent :

$$\begin{array}{lll} \text{Convergence monotone} & : & X_n \geq 0,\ X_n \uparrow X \ \Rightarrow\ E[X_n] \uparrow E[X]. \\ \text{Lemme de Fatou} & : & X_n \geq 0,\ \Rightarrow\ E[\liminf X_n] \leq \liminf E[X_n]. \\ \text{Convergence dominée} & : & |X_n| \leq Z\ E[Z] < \infty,\ X_n \longrightarrow X \text{ p.p.} \ \Rightarrow\ E[X_n] \longrightarrow E[X]. \end{array}$$

En théorie des probabilités on utilise l'expression presque sûrement (p.s. en abrégé) plutôt que le presque partout (p.p.) de la théorie de la mesure.

Les espaces $L^p(\Omega, \mathcal{A}, P)$ sont définis pour tout $p \in [1, \infty]$ comme dans le cours d'intégration. L'inégalité de Hölder s'écrit

$$E[|XY|] \leq E[|X|^p]^{1/p}\, E[|Y|^q]^{1/q},$$

pourvu que $\frac{1}{p} + \frac{1}{q} = 1$. En prenant $Y = 1$ on trouve $\|X\|_1 \leq \|X\|_p$, ce qui se généralise aussitôt à $\|X\|_r \leq \|X\|_p$ si $r \leq p$. En particulier $L^p(\Omega, \mathcal{A}, P) \subset L^r(\Omega, \mathcal{A}, P)$ si $r \leq p$.

L'inégalité de Cauchy-Schwarz s'écrit

$$E[|XY|] \leq E[X^2]^{1/2}\, E[Y^2]^{1/2}$$

et le cas particulier où $Y = 1$

$$E[|X|]^2 \leq E[X^2]$$

est très souvent utile.

Définition 8.2.1 *Soit $X \in L^2(\Omega, \mathcal{A}, P)$. La variance de X est*

$$\text{var}(X) = E[(X - E[X])^2]$$

et l'écart-type de X est

$$\sigma_X = \sqrt{\text{var}(X)}.$$

De manière informelle, $\text{var}(X)$ mesure la dispersion de X autour de sa moyenne $E[X]$. Remarquons que $\text{var}(X) = 0$ si et seulement si X est constante p.s.

Proposition 8.2.1 *On a aussi* $\text{var}(X) = E[X^2] - (E[X])^2$*, et pour tout* $a \in \mathbb{R}$*,*

$$E[(X-a)^2] = \text{var}(X) + (E[X]-a)^2.$$

En conséquence,

$$\text{var}(X) = \inf_{a\in\mathbb{R}} E[(X-a)^2].$$

Preuve. On a

$$E[(X-a)^2] = E[X^2] - 2a\,E[X] + a^2 = E[X^2] - (E[X])^2 + (E[X]-a)^2.$$

Les deux premières assertions en découlent aussitôt, en prenant $a = E[X]$ pour la première.
□

Inégalité de Markov. (cf cours d'intégration) Si $X \geq 0$ et $a > 0$,

$$P(X \geq a) \leq \frac{1}{a}\,E[X].$$

Inégalité de Bienaymé-Tchebicheff. Si $X \in L^2(\Omega, \mathcal{A}, P)$ et $a > 0$,

$$P(|X - E[X]| \geq a) \leq \frac{1}{a^2}\,\text{var}(X).$$

Cette inégalité découle de l'inégalité de Markov appliquée à la variable positive $(X - E[X])^2$.

Définition 8.2.2 *Soient* $X, Y \in L^2(\Omega, \mathcal{A}, P)$*. La covariance de* X *et* Y *est*

$$\text{cov}(X,Y) = E[(X-E[X])(Y-E[Y])] = E[X(Y-E[Y])] = E[XY] - E[X]E[Y].$$

Si $X = (X_1, \ldots, X_d)$ *est une variable aléatoire à valeurs dans* $\mathbb{R}^d$ *dont toutes les composantes sont dans* $L^2(\Omega, \mathcal{A}, P)$ *(ce qui équivaut à* $E[|X|^2] < \infty$*), la matrice de covariance de* X *est*

$$K_X = \Big(\text{cov}(X_i, X_j)\Big)_{1\leq i\leq d, 1\leq j\leq d}.$$

De manière informelle, la covariance de X et Y mesure la corrélation existant entre X et Y. Remarquons que $\text{cov}(X,X) = \text{var}(X)$ et que, d'après l'inégalité de Cauchy-Schwarz,

$$|\text{cov}(X,Y)| \leq \sqrt{\text{var}(X)}\,\sqrt{\text{var}(Y)}.$$

L'application $(X,Y) \longrightarrow \text{cov}(X,Y)$ est une forme bilinéaire sur $L^2(\Omega, \mathcal{A}, P)$.

Dans le cas vectoriel $X = (X_1, \ldots, X_d)$, la matrice K_X est symétrique positive : pour tous $\lambda_1, \ldots, \lambda_d \in \mathbb{R}^d$,

$$\sum_{i,j=1}^{d} \lambda_i\lambda_j\,K_X(i,j) = \text{var}\Big(\sum_{i=1}^{d} \lambda_i X_i\Big) \geq 0.$$

Exercice. Si A est une matrice (déterministe) $n \times d$ et $Y = AX$, vérifier que $K_Y = A\,K_X\,{}^tA$.

8.2.2 La régression linéaire

Soient $X, Y_1, \ldots, Y_n$ des variables aléatoires dans $L^2(\Omega, \mathcal{A}, P)$. On cherche à trouver la meilleure approximation de X comme fonction affine de $Y_1, \ldots, Y_n$. Précisément, on cherche à minimiser

$$E[(X - (\beta_0 + \beta_1 Y_1 + \cdots + \beta_n Y_n))^2]$$

sur tous les choix possibles du $(n+1)$-uplet de réels $(\beta_0, \ldots, \beta_n)$.

Proposition 8.2.2 *On a*

$$\inf_{\beta_0,\ldots,\beta_n \in \mathbb{R}} E[(X - (\beta_0 + \beta_1 Y_1 + \cdots + \beta_n Y_n))^2] = E[(X - Z)^2],$$

où

$$Z = E[X] + \sum_{j=1}^{n} \alpha_j (Y_j - E[Y_j]), \qquad (1)$$

les coefficients α_j étant (n'importe quelle) solution du système

$$\sum_{j=1}^{n} \alpha_j \operatorname{cov}(Y_j, Y_k) = \operatorname{cov}(X, Y_k) \ , \quad 1 \leq k \leq n.$$

En particulier, si K_Y est non-dégénérée, on a $\alpha = \operatorname{cov}(X, Y)\, K_Y^{-1}$ en notation matricielle.

Preuve. Soit H le sous-espace vectoriel de $L^2(\Omega, \mathcal{A}, P)$ engendré par $1, Y_1, \ldots, Y_n$. Alors, on sait que la variable aléatoire Z qui minimise $\|X - U\|_2$ pour $U \in H$ est la projection orthogonale de X sur H. On peut écrire Z sous la forme

$$Z = \alpha_0 + \sum_{j=1}^{n} \alpha_j (Y_j - E[Y_j]).$$

Par définition de la projection orthogonale, $X - Z$ est orthogonal à H. On doit donc avoir

$$E[(X - Z) \cdot 1] = 0,$$

d'où $\alpha_0 = E[X]$. De même, pour tout $k \in \{1, \ldots, n\}$,

$$E[(X - Z) \cdot (Y_k - E[Y_k])] = 0,$$

ce qui équivaut à $\operatorname{cov}(Z, Y_k) = \operatorname{cov}(X, Y_k)$, ou encore à

$$\sum_{j=1}^{n} \alpha_j \operatorname{cov}(Y_j, Y_k) = \operatorname{cov}(X, Y_k).$$

Inversement, si les coefficients α_j vérifient ce système d'équations, il est immédiat que la variable Z définie par le membre de droite de (1) est un élément de H tel que $X - Z$ soit orthogonal à H, donc doit coïncider avec la projection orthogonale de X sur H. □

Remarque. Si $n = 1$ et si on suppose que Y n'est pas constante p.s., on trouve que la meilleure (au sens L^2) approximation de X par une fonction affine de Y est

$$Z = E[X] + \frac{\operatorname{cov}(X, Y)}{\operatorname{var}(Y)} (Y - E[Y]).$$

C'est ce qu'on appelle parfois la droite de régression de X en Y.

8.2.3 Fonctions caractéristiques

Définition 8.2.3 *Si X est une variable aléatoire à valeurs dans $\mathbb{R}^d$, la fonction caractéristique de X est la fonction $\Phi_X : \mathbb{R}^d \longrightarrow \mathbb{C}$ définie par*

$$\Phi_X(\xi) = E[\exp(i\xi \cdot X)] \ , \qquad \xi \in \mathbb{R}^d.$$

On peut aussi écrire

$$\Phi_X(\xi) = \int e^{i\xi \cdot x} \, P_X(dx)$$

ce qui permet de voir Φ_X comme la transformée de Fourier de la loi de X. On écrit parfois $\Phi_X(\xi) = \widehat{P}_X(\xi)$. Le théorème de convergence dominée montre que Φ_X est continue (et bornée) sur $\mathbb{R}^d$.

Notre objectif est de montrer que la fonction caractéristique caractérise la loi de X. Nous commençons par un calcul important dans un cas particulier.

Lemme 8.2.3 *Soit X une variable aléatoire de loi gaussienne $\mathcal{N}(0, \sigma^2)$. Alors,*

$$\Phi_X(\xi) = \exp(-\frac{\sigma^2 \xi^2}{2}) \ , \qquad \xi \in \mathbb{R}.$$

Preuve. On a

$$\Phi_X(\xi) = \int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} \, e^{-x^2/(2\sigma^2)} e^{i\xi x} \, dx.$$

On se ramène facilement au cas $\sigma = 1$. Ensuite, un argument de parité montre que la partie imaginaire de $\Phi_X(\xi)$ est nulle. Il reste à calculer

$$f(\xi) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} \cos(\xi x) \, dx.$$

En dérivant sous le signe intégrale, on a

$$f'(\xi) = -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \, x \, e^{-x^2/2} \sin(\xi x) \, dx$$

(la justification est facile puisque $|x \, \sin(\xi x) \, e^{-x^2/2}| \leq |x| \, e^{-x^2/2}$ qui est intégrable). En intégrant par parties, il vient

$$f'(\xi) = -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} \, \xi \, \cos(\xi x) \, dx = -\xi \, f(\xi).$$

La fonction f est donc solution de l'équation différentielle $f'(\xi) = -\xi f(\xi)$, avec condition initiale $f(0) = 1$. Il en découle que $f(\xi) = \exp(-\xi^2/2)$. □

Théorème 8.2.4 *La fonction caractéristique d'une variable aléatoire X à valeurs dans $\mathbb{R}^d$ caractérise la loi de cette variable aléatoire. Autrement dit, la transformée de Fourier définie sur l'espace des mesures de probabilité sur $\mathbb{R}^d$ est injective.*

Preuve. On traite d'abord le cas $d = 1$. Pour tout $\sigma > 0$, soit g_σ la densité de la loi gaussienne $\mathcal{N}(0, \sigma^2)$:

$$g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{x^2}{2\sigma^2}) , \qquad x \in \mathbb{R}.$$

Si μ est une mesure de probabilité sur $\mathbb{R}$, on pose

$$\begin{aligned} f_\sigma(x) &= \int_{\mathbb{R}} g_\sigma(x-y)\,\mu(dy) \overset{\text{(def)}}{=} g_\sigma * \mu(x), \\ \mu_\sigma(dx) &= f_\sigma(x)\,dx. \end{aligned}$$

Pour montrer le résultat du théorème, il suffit d'établir que

1. μ_σ est déterminée par $\widehat{\mu}$.
2. Pour toute fonction $\varphi \in C_b(\mathbb{R})$, $\int \varphi(x)\mu_\sigma(dx) \longrightarrow \int \varphi(x)\mu(dx)$ quand $\sigma \to 0$.

Pour établir le point 1, on utilise le lemme pour écrire, pour tout $x \in \mathbb{R}$,

$$\sigma\sqrt{2\pi}\, g_\sigma(x) = \exp(-\frac{x^2}{2\sigma^2}) = \int_{\mathbb{R}} e^{i\xi x}\, g_{1/\sigma}(\xi)\, d\xi.$$

Il vient alors

$$\begin{aligned} f_\sigma(x) = \int_{\mathbb{R}} g_\sigma(x-y)\,\mu(dy) &= (\sigma\sqrt{2\pi})^{-1} \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} e^{i\xi(x-y)}\, g_{1/\sigma}(\xi)\, d\xi \Big) \mu(dy) \\ &= (\sigma\sqrt{2\pi})^{-1} \int_{\mathbb{R}} e^{i\xi x}\, g_{1/\sigma}(\xi) \Big(\int_R e^{-i\xi y} \mu(dy) \Big) d\xi \\ &= (\sigma\sqrt{2\pi})^{-1} \int_{\mathbb{R}} e^{i\xi x}\, g_{1/\sigma}(\xi)\, \widehat{\mu}(-\xi) d\xi. \end{aligned}$$

Dans l'avant-dernière égalité, on a utilisé le théorème de Fubini-Lebesgue, dont la justification est facile puisque μ est une mesure de probabilité et que la fonction $g_{1/\sigma}$ est intégrable pour la mesure de Lebesgue.

Pour le point 2, on écrit d'abord, pour toute fonction φ continue et bornée sur $\mathbb{R}$,

$$\int \varphi(x)\mu_\sigma(dx) = \int \varphi(x) \Big(\int g_\sigma(y-x)\mu(dy) \Big) dx = \int g_\sigma * \varphi(y)\mu(dy),$$

avec la même justification pour appliquer le théorème de Fubini-Lebesgue. Ensuite, on utilise les propriétés

$$\begin{aligned} &\int g_\sigma(x)\, dx = 1\,, \\ &\lim_{\sigma\to 0} \int_{\{|x|>\varepsilon\}} g_\sigma(x)\, dx = 0\,, \qquad \forall \varepsilon > 0, \end{aligned}$$

pour obtenir que, pour tout $y \in \mathbb{R}$,

$$\lim_{\varepsilon \to 0} g_\sigma * \varphi(y) = \varphi(y)$$

(cf les résultats du cours d'intégration concernant les approximations de la mesure de Dirac δ_0). Par convergence dominée, facile à justifier puisque $|g_\sigma * \varphi| \leq \sup|\varphi|$, on obtient

$$\lim_{\sigma\to 0}\int \varphi(x)\mu_\sigma(dx) = \int \varphi(x)\mu(dx),$$

ce qui termine la preuve dans le cas $d = 1$.

La preuve dans le cas d quelconque est similaire. On utilise les fonctions

$$g_\sigma^{(d)}(x_1,\ldots,x_d) = \prod_{j=1}^{d} g_\sigma(x_j)$$

en remarquant que pour $\xi \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} g_\sigma^{(d)}(x)\, e^{i\xi\cdot x}\, dx = \prod_{j=1}^{d}\int g_\sigma(x_j)\, e^{i\xi_j\cdot x_j}\, dx_j = (2\pi\sigma)^{d/2} g_{1/\sigma}^{(d)}(\xi).$$

□

Proposition 8.2.5 *Soit $X = (X_1,\ldots,X_d)$ une v.a. à valeurs dans $\mathbb{R}^d$ et de carré intégrable. Alors Φ_X est de classe C^2 et*

$$\Phi_X(\xi) = 1 + i\sum_{j=1}^{d}\xi_j E[X_j] - \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d}\xi_j\xi_k E[X_jX_k] + o(|\xi|^2)$$

quand $\xi = (\xi_1,\ldots,\xi_d)$ tend vers 0.

Preuve. En dérivant sous le signe intégrale, on trouve

$$\frac{\partial\Phi_X}{\partial\xi_j}(\xi) = i\, E[X_j e^{i\xi\cdot X}],$$

la justification étant facile puisque $|iX_je^{i\xi\cdot X}| = |X_j|$ et $X_j \in L^2 \subset L^1$. De même, puisque $E[|X_jX_k|] \leq E[X_j^2]^{1/2}E[X_k^2]^{1/2} < \infty$, on peut dériver une seconde fois et trouver que

$$\frac{\partial^2\Phi_X}{\partial\xi_j\partial\xi_k}(\xi) = -\,E[X_jX_ke^{i\xi\cdot X}].$$

De plus le théorème de continuité sous le signe intégrale assure que $\frac{\partial^2\Phi_X}{\partial\xi_j\partial\xi_k}(\xi)$ est fonction continue de ξ.

Enfin la dernière assertion est simplement le développement de Taylor de Φ_X à l'ordre 2 à l'origine. □

Remarque. Si on suppose que X est de puissance p-ième intégrable ($p \geq 1$ entier) le même raisonnement montre que Φ_X est de classe C^p. C'est cependant le cas $p = 2$ qui sera le plus utile dans la suite.

8.2.4 Fonction génératrice

Dans le cas de variables aléatoires à valeurs dans $\mathbb{N}$, on utilise les fonctions génératrices plutôt que les fonctions caractéristiques.

Définition 8.2.4 *Soit X une v.a. à valeurs dans $\mathbb{N}$. La fonction génératrice de X est la fonction g_X définie sur l'intervalle $[0,1]$ par*

$$g_X(r) = E[r^X] = \sum_{n=0}^{\infty} P(X=n)\, r^n.$$

La fonction g_X est continue sur $[0,1]$ (cela découle par exemple du théorème de convergence dominée), et on a $g_X(0) = P(X=0)$ et $g_X(1) = 1$. Le rayon de convergence de la série entière qui apparaît dans la définition est donc supérieur ou égal à un. Cela montre que la fonction génératrice g_X caractérise la loi de X, puisque les nombres $P(X=n)$ apparaissent comme les coefficients du développement de Taylor de g_X en 0.

On voit facilement que g_X a toujours une dérivée à gauche en 1, éventuellement infinie, et que

$$g'_X(1) = E[X].$$

Plus généralement, pour tout entier $p \geq 1$,

$$\lim_{r \uparrow 1} g_X^{(p)}(r) = E[X(X-1)\cdots(X-p+1)]$$

ce qui montre comment retrouver tous les moments de X à partir de la connaissance de la fonction génératrice.

Chapitre 9

Indépendance

Le concept d'indépendance est sans doute la première notion importante où la théorie des probabilités se différencie nettement de l'intégration. S'il est plus facile de comprendre intuitivement la définition de l'indépendance de deux événements ou de deux variables aléatoires, la notion la plus fondamentale est celle de l'indépendance de deux (ou plusieurs) sous-tribus. Un résultat-clé de ce chapitre relie l'indépendance de deux variables aléatoires au fait que la loi du couple formé par ces deux variables est la mesure-produit des lois individuelles. Avec le théorème de Fubini, cela permet des reformulations souvent utiles de la notion d'indépendance. A titre d'application, on établit le célèbre lemme de Borel-Cantelli (dont une application amusante donne des propriétés surprenantes du développement dyadique d'un nombre réel choisi au hasard) et une première forme de la loi des grands nombres, qui suffit à établir le lien entre notre approche axiomatique des probabilités et la définition "historique" (probabilité d'un événement = fréquence d'apparition de cet événement lorsqu'on répète un grand nombre de fois la même expérience aléatoire).

9.1 Evénements indépendants

Dans tout ce chapitre on se place sur un espace de probabilité $(\Omega, \mathcal{A}, P)$. Si $A, B \in \mathcal{A}$ sont deux événements, on dit que A et B sont indépendants si

$$P(A \cap B) = P(A)P(B).$$

Au moins lorsque $P(B) > 0$, on peut interprèter cette définition en disant que la probabilité conditionnelle

$$P(A \mid B) \overset{\text{(def)}}{=} \frac{P(A \cap B)}{P(B)}$$

coïncide avec $P(A)$: le fait de savoir que B est réalisé ne donne pas d'information sur la réalisation ou non de l'événement A (et on peut intervertir les rôles de A et B).

Exemples. (i) *Lancer de deux dés* : $\Omega = \{1, 2, \dots, 6\}^2$, $P(\{\omega\}) = 1/36$ pour tout $\omega \in \Omega$. Les événements $A = \{6\} \times \{1, 2, \dots, 6\}$ et $B = \{1, 2, \dots, 6\} \times \{6\}$ sont indépendants. En fait la probabilité P a été construite précisément pour qu'un événement relatif au résultat du premier lancer soit indépendant d'un événement relatif au résultat du second.

(ii) *Lancer d'un seul dé* : $\Omega = \{1, 2, \ldots, 6\}$, $P(\{\omega\}) = 1/6$ pour tout $\omega \in \Omega$. Les événements $A = \{1, 2\}$ et $B = \{1, 3, 5\}$ sont indépendants.

Définition 9.1.1 *On dit que n événements $A_1, \ldots, A_n$ sont indépendants si, pour tout sous-ensemble non vide $\{j_1, \ldots, j_p\}$ de $\{1, \ldots, n\}$, on a*

$$P(A_{j_1} \cap A_{j_2} \cap \ldots \cap A_{j_p}) = P(A_{j_1})\, P(A_{j_2}) \ldots P(A_{j_p}).$$

Remarques. Il ne suffit pas que l'on ait

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\, P(A_2) \ldots P(A_n).$$

Il ne suffit pas non plus que, pour chaque paire $\{i, j\} \subset \{1, \ldots, n\}$, les événements A_i et A_j soient indépendants. Pour donner un exemple, considérons l'espace correspondant à deux lancers de pile ou face (pièce non truquée) et prenons

$$\begin{aligned} A &= \{\text{pile au premier lancer}\} \\ B &= \{\text{pile au second lancer}\} \\ C &= \{\text{même résultat aux deux lancers}\}. \end{aligned}$$

Les événements A, B, C sont indépendants deux à deux mais non indépendants.

Proposition 9.1.1 *Les n événements $A_1, \ldots, A_n$ sont indépendants si et seulement si on a*

$$P(B_1 \cap \ldots \cap B_n) = P(B_1) \ldots P(B_n)$$

dès que $B_i \in \sigma(A_i) = \{\varnothing, A_i, A_i^c, \Omega\}$ pour tout $i \in \{1, \ldots, n\}$.

Preuve. Il est clair que la condition donnée est plus forte que celle de la définition : prendre $B_i = A_i$ si $i \in \{j_1, \ldots, j_p\}$ et $B_i = \Omega$ sinon. Inversement, supposons que $A_1, \ldots, A_n$ sont indépendants. Pour vérifier la propriété de la proposition, on peut supposer $B_i \neq \varnothing$ pour tout $i \in \{1, \ldots, n\}$. Ensuite, si $\{j_1, \ldots, j_p\} = \{i : B_i \neq \Omega\}$, on est ramené à montrer que

$$P(B_{j_1} \cap B_{j_2} \cap \ldots \cap B_{j_p}) = P(B_{j_1})\, P(B_{j_2}) \ldots P(B_{j_p}),$$

dès que $B_{j_k} = A_{j_k}$ ou $A_{j_k}^c$. Finalement, il suffit de montrer que si $C_1, C_2, \ldots, C_p$ sont indépendants, $C_1^c, C_2, \ldots, C_p$ le sont aussi. Mais cela est facile puisque, pour tout sous-ensemble $\{i_1, \ldots, i_q\}$ de $\{2, \ldots, p\}$,

$$\begin{aligned} P(C_1^c \cap C_{i_1} \cap \cdots \cap C_{i_q}) &= P(C_{i_1} \cap \cdots \cap C_{i_q}) - P(C_1 \cap C_{i_1} \cap \cdots \cap C_{i_q}) \\ &= P(C_{i_1}) \ldots P(C_{i_q}) - P(C_1) P(C_{i_1}) \ldots P(C_{i_q}) \\ &= P(C_1^c) P(C_{i_1}) \ldots P(C_{i_q}) \end{aligned}$$

□

9.2 Variables aléatoires et tribus indépendantes

La notion la plus générale est celle de tribus indépendantes.

Définition 9.2.1 *Soient $\mathcal{B}_1, \ldots, \mathcal{B}_n$ n sous-tribus de $\mathcal{A}$. On dit que $\mathcal{B}_1, \ldots, \mathcal{B}_n$ sont indépendantes si et seulement si*

$$\forall A_1 \in \mathcal{B}_1, \ldots, \forall A_n \in \mathcal{B}_n, \qquad P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\, P(A_2) \ldots P(A_n).$$

Soient $X_1, \ldots, X_n$ n variables aléatoires à valeurs respectivement dans $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$. On dit que les variables $X_1, \ldots, X_n$ sont indépendantes si les tribus $\sigma(X_1), \ldots, \sigma(X_n)$ le sont. Cela équivaut encore à dire que

$$\forall F_1 \in \mathcal{E}_1, \ldots, \forall F_n \in \mathcal{E}_n,\; P(\{X_1 \in F_1\} \cap \ldots \cap \{X_n \in F_n\}) = P(X_1 \in F_1) \ldots P(X_n \in F_n) \tag{9.1}$$

(en effet on sait que $\sigma(X_i) = \{X_i^{-1}(F) : F \in \mathcal{E}_i\}$).

De manière intuitive, les v.a. $X_1, \ldots, X_n$ sont indépendantes si la connaissance de certaines d'entre elles ne donne pas d'information sur les autres.

Remarques. (i) Si $\mathcal{B}_1, \ldots, \mathcal{B}_n$ sont n sous-tribus indépendantes, et si, pour tout $i \in \{1, \ldots, n\}$, X_i est une v.a. $\mathcal{B}_i$-mesurable, alors $X_1, \ldots, X_n$ sont indépendantes.

(ii) Les n événements $A_1, \ldots, A_n$ sont indépendants si et seulement si les tribus $\sigma(A_1), \ldots,$ $\sigma(A_n)$ le sont (cf proposition précédente).

Si $X_1, \ldots, X_n$ sont des variables aléatoires à valeurs dans $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$ respectivement, le n-uplet $(X_1, \ldots, X_n)$ est une v.a. à valeurs dans l'espace $E_1 \times \cdots \times E_n$ muni de la tribu produit $\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n$.

Théorème 9.2.1 *Les n variables aléatoires $X_1, \ldots, X_n$ sont indépendantes si et seulement si la loi du n-uplet $(X_1, \ldots, X_n)$ est le produit des lois de $X_1, \ldots, X_n$:*

$$P_{(X_1,\ldots,X_n)} = P_{X_1} \otimes \cdots \otimes P_{X_n}.$$

De plus, on a alors

$$E\Big[\prod_{i=1}^{n} f_i(X_i)\Big] = \prod_{i=1}^{n} E[f_i(X_i)]$$

dès que f_i est une fonction mesurable positive sur $(E_i, \mathcal{E}_i)$, pour tout $i \in \{1, \ldots, n\}$.

Preuve. Soit $F_i \in \mathcal{E}_i$, pour tout $i \in \{1, \ldots, n\}$. On a d'une part

$$P_{(X_1,\ldots,X_n)}(F_1 \times \cdots \times F_n) = P(\{X_1 \in F_1\} \cap \ldots \cap \{X_n \in F_n\})$$

et d'autre part

$$P_{X_1} \otimes \cdots \otimes P_{X_n}(F_1 \times \cdots \times F_n) = \prod_{i=1}^{n} P_{X_i}(F_i) = \prod_{i=1}^{n} P(X_i \in F_i).$$

En comparant avec (9.1), on voit que $X_1, \dots, X_n$ sont indépendantes si et seulement si les deux mesures de probabilité $P_{(X_1,\dots,X_n)}$ et $P_{X_1} \otimes \cdots \otimes P_{X_n}$ prennent les mêmes valeurs sur les pavés $F_1 \times \cdots \times F_n$. Mais comme on sait (lemme de classe monotone) qu'une mesure de probabilité sur un espace-produit est caractérisée par ses valeurs sur les pavés, cela équivaut encore à dire que $P_{(X_1,\dots,X_n)} = P_{X_1} \otimes \cdots \otimes P_{X_n}$.

La deuxième assertion est ensuite une conséquence du théorème de Fubini-Tonnelli :

$$\begin{aligned} E\Big[\prod_{i=1}^n f_i(X_i)\Big] &= \int_{E_1\times\cdots\times E_n} \prod_{i=1}^n f_i(x_i)\, P_{X_1}(dx_1)\dots P_{X_n}(dx_n) \\ &= \prod_{i=1}^n \int_{E_i} f_i(x_i)\, P_{X_i}(dx_i) \\ &= \prod_{i=1}^n E[f_i(X_i)]. \end{aligned}$$

□

Le théorème ci-dessus montre aussi comment construire des v.a. indépendantes. Considérons le cas de v.a. réelles, et soient $\mu_1, \dots, \mu_n$ des mesures de probabilité sur $\mathbb{R}^n$. Alors, comme on l'a observé dans le Chapitre 8, on peut construire une v.a. $Y = (Y_1, \dots, Y_n)$ à valeurs dans $\mathbb{R}^n$ dont la loi est $\mu_1 \otimes \cdots \otimes \mu_n$. D'après le théorème précédent, les composantes $Y_1, \dots Y_n$ de Y sont des v.a. réelles indépendantes de lois respectives $\mu_1, \dots, \mu_n$.

Remarques. Si les fonctions f_i sont de signe quelconque, l'égalité

$$E\Big[\prod_{i=1}^n f_i(X_i)\Big] = \prod_{i=1}^n E[f_i(X_i)]$$

reste vraie à condition que $E[|f_i(X_i)|] < \infty$ pour tout $i \in \{1, \dots, n\}$, et on a alors aussi

$$E\Big[\prod_{i=1}^n |f_i(X_i)|\Big] = \prod_{i=1}^n E[|f_i(X_i)|] < \infty$$

ce qui justifie l'existence du terme de gauche dans la formule précédente.

En particulier, si $X_1, \dots, X_n$ sont n v.a. réelles indépendantes et dans L^1, on a aussi $X_1 \cdots X_n \in L^1$, et

$$E[X_1 \cdots X_n] = \prod_{i=1}^n E[X_i].$$

Remarquons qu'en général le produit de v.a. dans L^1 n'est pas dans L^1 (l'indépendance est une propriété très particulière).

Corollaire 9.2.2 *Si X_1, X_2 sont deux variables aléatoires réelles indépendantes et dans L^2, on a* $\mathrm{cov}(X_1, X_2) = 0$.

Cela découle de ce qui précède puisque $\mathrm{cov}(X_1, X_2) = E[X_1X_2] - E[X_1]E[X_2]$.

La réciproque du corollaire est fausse. La propriété de covariance nulle (pour deux v.a. dans L^2) est beaucoup plus faible que l'indépendance. Pour donner un exemple, partons d'une v.a. réelle X_1 dont la loi a une densité notée $p(x)$ symétrique ($p(x) = p(-x)$) et telle que $\int x^2 p(x)dx < \infty$ (de sorte que $X_1 \in L^2$). On peut par exemple choisir pour X_1 une v.a. de loi $\mathcal{N}(0, \sigma^2)$. Soit ensuite ε une deuxième v.a. à valeurs dans $\{-1, 1\}$, indépendante de X_1 et telle que $P(\varepsilon = 1) = P(\varepsilon = -1) = \frac{1}{2}$. Alors, si $X_2 = \varepsilon X_1$, on voit immédiatement que $\text{cov}(X_1, X_2) = 0$ alors que X_1 et X_2 ne sont pas indépendantes. En effet, si X_1 et X_2 l'étaient, $|X_1|$ serait indépendante de $|X_2| = |X_1|$. Or si une v.a. réelle est indépendante d'elle-même, elle doit être constante p.s. (exercice !) et donc sa loi est une mesure de Dirac. C'est une contradiction puisque la loi de $|X_1|$ a une densité donnée par $2\,p(x)1_{\mathbb{R}_+}(x)$.

Corollaire 9.2.3 *Soient $X_1, \ldots, X_n$ n variables aléatoires réelles.*
(i) *Supposons d'abord que, pour tout $i \in \{1, \ldots, n\}$, la loi de X_i a une densité notée p_i, et que les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes. Alors, la loi de $(X_1, \ldots, X_n)$ a une densité donnée par*

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_i(x_i).$$

(ii) *Inversement, supposons que la loi de $(X_1, \ldots, X_n)$ a une densité de la forme*

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} q_i(x_i),$$

où les fonctions q_i sont boréliennes positives sur $\mathbb{R}$. Alors les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes et pour chaque $i \in \{1, \ldots, n\}$, la loi de X_i a une densité p_i qui s'écrit $p_i = C_i\, q_i$, où $C_i > 0$ est une constante.

Preuve. La première partie est une conséquence immédiate du théorème ci-dessus, puisque si $P_{X_i}(dx_i) = p_i(x_i)dx_i$, le théorème de Fubini-Tonnelli montre que

$$P_{X_1} \otimes \cdots \otimes P_{X_n}(dx_1 \ldots dx_n) = \Big(\prod_{i=1}^{n} p_i(x_i)\Big) dx_1 \ldots dx_n.$$

Pour la partie (ii), on remarque d'abord que, toujours à l'aide du théorème de Fubini-Tonnelli, on a

$$\prod_{i=1}^{n} \Big(\int q_i(x)dx\Big) = \int_{\mathbb{R}^n} p(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1,$$

et en particulier $K_i := \int q_i(x)dx) \in]0, \infty[$ pour tout $i \in \{1, \ldots, n\}$. Ensuite, d'après un résultat du Chapitre 8, la densité de X_i est

$$p_i(x_i) = \int_{\mathbb{R}^{n-1}} p(x_1, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots, dx_n = \Big(\prod_{j \neq i} K_j\Big) q_i(x_i) = \frac{1}{K_i} q_i(x_i).$$

Cela permet de réécrire la densité de $(X_1, \ldots, X_n)$ sous la forme

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} q_i(x_i) = \prod_{i=1}^{n} p_i(x_i)$$

et on voit que $P_{(X_1,\ldots,X_n)} = P_{X_1} \otimes \cdots \otimes P_{X_n}$ d'où l'indépendance. $\square$

Exemple. Soit U une variable de loi exponentielle de paramètre 1 et soit V une variable uniforme sur l'intervalle $[0,1]$. On suppose que U et V sont indépendantes. Alors, si on définit

$$X = \sqrt{U}\,\cos(2\pi V)\ ,\quad Y = \sqrt{U}\,\sin(2\pi V),$$

les deux variables aléatoires X et Y sont indépendantes. Pour le voir calculons la loi du couple (X,Y). Pour toute fonction φ mesurable positive sur $\mathbb{R}^2$,

$$\begin{aligned} E[\varphi(X,Y)] &= \int_0^\infty \int_0^1 \varphi(\sqrt{u}\cos(2\pi v), \sqrt{u}\sin(2\pi v))\, e^{-u}\, du dv \\ &= \frac{1}{\pi}\int_0^\infty \int_0^{2\pi} \varphi(r\cos\theta, r\sin\theta)\, re^{-r^2}\, drd\theta \\ &= \frac{1}{\pi}\int_{\mathbb{R}^2} \varphi(x,y)\, e^{-x^2-y^2}\, dxdy. \end{aligned}$$

On obtient que la loi du couple (X,Y) a pour densité $\pi^{-1}\exp(-x^2-y^2)$ qui a une forme produit comme dans la partie (ii) de la proposition. Donc X et Y sont indépendantes (on voit aussi que X et Y ont la même densité

$$p(x) = \frac{1}{\sqrt{\pi}}\exp(-x^2)$$

et donc X et Y suivent chacune la loi $\mathcal{N}(0,1/2)$).

Remarque. Si $X_1,\ldots,X_n$ sont n variables aléatoires réelles, il y a équivalence entre :

(i) $X_1,\ldots,X_n$ sont indépendantes.

(ii) Pour tous $a_1,\ldots,a_n \in \mathbb{R}$, $P(X_1 \leq a_1,\ldots,X_n \leq a_n) = \prod_{i=1}^n P(X_i \leq a_i)$.

(iii) Si $f_1,\ldots,f_n$ sont continues à support compact de $\mathbb{R}$ dans $\mathbb{R}_+$,

$$E\Big[\prod_{i=1}^n f_i(X_i)\Big] = \prod_{i=1}^n E[f_i(X_i)].$$

(iv) La fonction caractéristique de X est

$$\Phi_X(\xi_1,\ldots,\xi_n) = \prod_{i=1}^n \Phi_{X_i}(\xi_i)$$

(pour montrer (iv)$\Rightarrow$(i), utiliser l'injectivité de la transformée de Fourier, cf Chapitre 8).

Nous passons maintenant à un résultat technique très utile.

Proposition 9.2.4 *Soient $\mathcal{B}_1, \ldots, \mathcal{B}_n$ des sous-tribus de $\mathcal{A}$. Pour tout $i \in \{1, \ldots, n\}$, soit $\mathcal{C}_i \subset \mathcal{B}_i$ une classe stable par intersections finies, contenant Ω et telle que $\sigma(\mathcal{C}_i) = \mathcal{B}_i$. Supposons que*

$$\forall C_1 \in \mathcal{C}_1, \ldots, \forall C_n \in \mathcal{C}_n, \qquad P(C_1 \cap C_2 \cap \ldots \cap C_n) = P(C_1)\, P(C_2) \ldots P(C_n).$$

Alors, les tribus $\mathcal{B}_1, \ldots, \mathcal{B}_n$ sont indépendantes.

Preuve. Fixons d'abord $C_2 \in \mathcal{C}_2, \ldots, C_n \in \mathcal{C}_n$, et posons

$$\mathcal{M}_1 = \{B_1 \in \mathcal{B}_1 : P(B_1 \cap C_2 \cap \ldots \cap C_n) = P(B_1)\, P(C_2) \ldots P(C_n)\}.$$

Alors $\mathcal{C}_1 \subset \mathcal{M}_1$ par hypothèse, et d'autre part on voit facilement que $\mathcal{M}_1$ est une classe monotone. Le lemme de classe monotone entraîne que $\mathcal{M}_1$ contient $\sigma(\mathcal{C}_1) = \mathcal{B}_1$, et on a montré

$$\forall B_1 \in \mathcal{B}_1, \forall C_2 \in \mathcal{C}_2, \ldots, \forall C_n \in \mathcal{C}_n, \qquad P(B_1 \cap C_2 \cap \ldots \cap C_n) = P(B_1)\, P(C_2) \ldots P(C_n).$$

Pour continuer, on fixe $B_1 \in \mathcal{B}_1, C_3 \in \mathcal{C}_3, \ldots, C_n \in \mathcal{C}_n$ et on pose

$$\mathcal{M}_1 = \{B_2 \in \mathcal{B}_2 : P(B_1 \cap B_2 \cap C_3 \cap \ldots \cap C_n) = P(B_1)\, P(B_2)\, P(C_3) \ldots P(C_n)\}.$$

A nouveau, $\mathcal{M}_2$ est une classe monotone qui contient $\mathcal{C}_2$ et donc aussi $\sigma(\mathcal{C}_2) = \mathcal{B}_2$. En raisonnant par récurrence, on arrive facilement au résultat voulu. □

Conséquence. Regroupement par paquets. Soient $\mathcal{B}_1, \ldots, \mathcal{B}_n$ des tribus indépendantes, et soient $n_0 = 0 < n_1 < \cdots < n_p = n$. Alors les tribus

$$\begin{aligned}
\mathcal{D}_1 &= \mathcal{B}_1 \vee \cdots \vee \mathcal{B}_{n_1} \overset{\text{(not)}}{=} \sigma(\mathcal{B}_1, \ldots, \mathcal{B}_{n_1}) \\
\mathcal{D}_2 &= \mathcal{B}_{n_1+1} \vee \cdots \vee \mathcal{B}_{n_2} \\
&\cdots \\
\mathcal{D}_p &= \mathcal{B}_{n_{p-1}+1} \vee \cdots \vee \mathcal{B}_{n_p}
\end{aligned}$$

sont indépendantes. Pour le voir, il suffit d'appliquer la proposition ci-dessus en prenant pour $\mathcal{C}_j$ la classe des parties de la forme

$$B_{n_{j-1}+1} \cap \cdots \cap B_{n_j}$$

où $B_i \in \mathcal{B}_i$ pour tout $i \in \{n_{j-1} + 1, \ldots, n_j\}$.

En particulier, si $X_1, \ldots, X_n$ sont indépendantes, les v.a.

$$Y_1 = (X_1, \ldots, X_{n_1}), \ldots, Y_p = (X_{n_{p-1}+1}, \ldots, X_{n_p})$$

sont indépendantes.

Exemple. Si $X_1, \ldots, X_4$ sont des v.a. réelles indépendantes, les v.a.

$$Z_1 = X_1 X_3 \ , \qquad Z_2 = X_2^3 + X_4$$

sont indépendantes.

Indépendance d'une famille infinie. Soit $(\mathcal{B}_i)_{i\in I}$ une famille quelconque de sous-tribus de $\mathcal{A}$. On dit que cette famille est indépendante si pour tout sous-ensemble fini $\{i_1,\ldots,i_p\}$ de I, les tribus $\mathcal{B}_{i_1},\ldots,\mathcal{B}_{i_p}$ sont indépendantes.

Si $(X_i)_{i\in I}$ est une famille quelconque de variables aléatoires, cette famille est dite indépendante si la famille de tribus $(\sigma(X_i))_{i\in I}$ l'est.

Proposition 9.2.5 *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes. Alors, pour tout entier $p\in\mathbb{N}$, les deux tribus*

$$\mathcal{B}_1=\sigma(X_0,\ldots,X_p)\ ,\qquad \mathcal{B}_2=\sigma(X_{p+1},X_{p+2},\ldots)$$

sont indépendantes.

Preuve. Il suffit d'appliquer la proposition précédente en prenant

$$\begin{aligned}\mathcal{C}_1&=\sigma(X_0,\ldots,X_p)=\mathcal{B}_1\\ \mathcal{C}_2&=\bigcup_{k=p+1}^{\infty}\sigma(X_{p+1},X_{p+2},\ldots,X_k)\subset\mathcal{B}_2\end{aligned}$$

et en remarquant que l'hypothèse est satisfaite grâce au principe du regroupement par paquets. □

9.3 Le lemme de Borel-Cantelli

Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements on note

$$\limsup A_n=\bigcap_{n=0}^{\infty}\Big(\bigcup_{k=n}^{\infty}A_k\Big)$$

et

$$\liminf A_n=\bigcup_{n=0}^{\infty}\Big(\bigcap_{k=n}^{\infty}A_k\Big)$$

Lemme 9.3.1 *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements.*

(i) *Si $\sum_{n\in\mathbb{N}}P(A_n)<\infty$, alors*

$$P(\limsup A_n)=0$$

ou de manière équivalente,

$$p.s.\qquad \{n\in\mathbb{N}:\omega\in A_n\}\ \textit{est fini.}$$

(ii) *Si* $\sum_{n\in\mathbb{N}} P(A_n) = \infty$ *et si les événements* A_n *sont indépendants, alors*

$$P(\limsup A_n) = 1$$

ou de manière équivalente,

$$p.s. \qquad \{n \in \mathbb{N} : \omega \in A_n\} \text{ est infini.}$$

Remarque. L'hypothèse d'indépendance (ou une autre hypothèse convenable) est nécessaire dans (ii), comme le montre l'exemple trivial où $A_n = A$ pour tout $n \in \mathbb{N}$, avec $0 < P(A) < 1$.

Preuve. (i) Si $\sum_{n\in\mathbb{N}} P(A_n) < \infty$, alors

$$E\Big[\sum_{n\in\mathbb{N}} 1_{A_n}\Big] = \sum_{n\in\mathbb{N}} P(A_n) < \infty$$

et donc $\sum_{n\in\mathbb{N}} 1_{A_n} < \infty$ p.s.
(ii) Fixons d'abord $n_0 \in \mathbb{N}$, et observons que si $n \geq n_0$,

$$P\Big(\bigcap_{k=n_0}^{n} A_k^c\Big) = \prod_{k=n_0}^{n} P(A_k^c) = \prod_{k=n_0}^{n} (1 - P(A_k)).$$

La divergence de la série $\sum P(A_k)$ entraîne alors que

$$P\Big(\bigcap_{k=n_0}^{\infty} A_k^c\Big) = 0.$$

Comme cela est vrai pour tout $n_0 \in \mathbb{N}$, on a aussi

$$P\Big(\bigcup_{n_0=0}^{\infty} \Big(\bigcap_{k=n_0}^{\infty} A_k^c\Big)\Big) = 0$$

et, en passant au complémentaire,

$$P\Big(\bigcap_{n_0=0}^{\infty} \Big(\bigcup_{k=n_0}^{\infty} A_k\Big)\Big) = 1,$$

ce qui est le résultat voulu. □

Deux applications. (1) Il n'existe pas de mesure de probabilité sur $\mathbb{N}$ telle que la probabilité de l'ensemble des multiples de n soit égale à $1/n$ pour tout entier $n \geq 1$. En effet, supposons qu'il existe une telle probabilité, notée P. Soit $\mathcal{P}$ l'ensemble des nombres premiers et pour tout $p \in \mathcal{P}$, notons $A_p = p\mathbb{N}$ l'ensemble des multiples de p. Alors, il est facile de voir que les A_p, $p \in \mathcal{P}$, sont indépendants. En effet, si $p_1, \ldots, p_k$ sont des nombres premiers distincts,

$$P(A_{p_1} \cap \ldots \cap A_{p_k}) = P(p_1\mathbb{N} \cap \ldots \cap p_k\mathbb{N}) = P((p_1 \ldots p_k)\mathbb{N}) = \frac{1}{p_1 \ldots p_k} = \prod_{j=1}^{k} P(A_{p_j}).$$

Par ailleurs, on sait que

$$\sum_{p\in\mathcal{P}} P(A_p) = \sum_{p\in\mathcal{P}} \frac{1}{p} = \infty.$$

On peut donc appliquer la partie (ii) du lemme de Borel-Cantelli pour obtenir que presque tout (au sens de la probabilité P) entier n appartient à une infinité d'ensembles A_p, et donc est multiple d'une infinité de nombres premiers distincts. C'est évidemment absurde.

(2) Considérons le cas où

$$(\Omega, \mathcal{A}, P) = ([0,1[, \mathcal{B}([0,1[), \lambda).$$

Pour tout $n \geq 1$, on pose

$$\forall \omega \in [0,1[, \quad X_n(\omega) = [2^n\omega] - 2[2^{n-1}\omega],$$

où $[x]$ désigne la partie entière d'un nombre réel x. Alors $X_n(\omega) \in \{0,1\}$ et on vérifie aisément par récurrence sur n que, pour tout $\omega \in [0,1[$,

$$0 \leq \omega - \sum_{k=1}^{n} X_k(\omega) 2^{-k} < 2^{-n},$$

ce qui montre que

$$\omega = \sum_{k=1}^{\infty} X_k(\omega)\, 2^{-k}.$$

Les nombres $X_k(\omega)$ sont donc les coefficients du développement dyadique (propre) de ω. En explicitant l'ensemble $\{X_n = 1\}$ on montre facilement que pour tout $n \geq 1$,

$$P(X_n = 0) = P(X_n = 1) = \frac{1}{2}.$$

Enfin, on observe que la suite $(X_n)_{n\geq 1}$ est indépendante. En effet, il suffit ici de vérifier que, pour tous $i_1, \ldots, i_p \in \{0,1\}$, on a

$$P(X_1 = i_1, \ldots, X_p = i_p) = \frac{1}{2^p} = \prod_{j=1}^{p} P(X_j = i_j).$$

Or, on voit immédiatement que

$$\{X_1 = i_1, \ldots, X_p = i_p\} = [\sum_{j=1}^{p} i_j 2^{-j}, \sum_{j=1}^{p} i_j 2^{-j} + 2^{-p}[,$$

d'où le résultat voulu.

Soit $p \geq 1$ un entier quelconque, et soient $i_1, \ldots, i_p \in \{0,1\}$. Alors, le lemme de Borel-Cantelli permet de voir que

$$\text{p.s. } \operatorname{Card}\{k \geq 0 : X_{k+1} = i_1, \ldots, X_{k+p} = i_p\} = \infty. \tag{9.2}$$

Cela montre qu'une suite finie donnée de 0 et de 1 apparaît une infinité de fois dans le développement dyadique de presque tout (au sens de la mesure de Lebesgue) réel de l'intervalle $[0,1[$. Pour établir (9.2), il suffit de poser, pour tout entier $n \in \mathbb{N}$,

$$Y_n = (X_{np+1}, X_{np+2}, \ldots, X_{np+p}).$$

Le principe du regroupement par paquets montre que la suite $(Y_n)_{n\in\mathbb{N}}$ est indépendante, et le résultat recherché découle d'une application du lemme de Borel-Cantelli à la suite des événements

$$A_n = \{Y_n = (i_1, \ldots, i_p)\}$$

qui sont indépendants et tous de probabilité 2^{-p}.

Puisqu'une réunion dénombrable d'ensembles de probabilité nulle est encore de probabilité nulle, on peut renforcer (9.2) sous la forme

$$\text{p.s. } \forall p \geq 1, \quad \forall i_1, \ldots, i_p \in \{0,1\}, \quad \text{Card}\{k \geq 0 : X_{k+1} = i_1, \ldots, X_{k+p} = i_p\} = \infty.$$

Autrement dit, pour presque tout réel x de $[0,1[$, n'importe quelle suite finie de 0 et de 1 apparaît une infinité de fois dans le développement dyadique de x.

9.4 Sommes de variables aléatoires indépendantes.

Les sommes de variables aléatoires indépendantes jouent un rôle important en théorie des probabilités, et seront étudiées dans le chapitre suivant. Nous regroupons d'abord quelques propriétés importantes sous la forme d'une proposition. Si μ et ν sont deux mesures de probabilité sur $\mathbb{R}^d$, on note $\mu * \nu$ la mesure-image de $\mu \otimes \nu$ par l'application $(x,y) \longrightarrow x+y$: pour toute fonction mesurable positive φ sur $\mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \varphi(z)\, \mu * \nu(dz) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(x+y)\, \mu(dx)\nu(dy).$$

Proposition 9.4.1 *Soient X et Y deux variables aléatoires indépendantes à valeurs dans $\mathbb{R}^d$.*
(i) *La loi de $X+Y$ est $P_X * P_Y$. En particulier, si X a une densité notée p_X et Y a une densité notée p_Y, $X+Y$ a pour densité $p_X * p_Y$.*
(ii) *La fonction caractéristique de $X+Y$ est $\Phi_{X+Y}(\xi) = \Phi_X(\xi)\Phi_Y(\xi)$. (De manière équivalente, si μ et ν sont deux mesures de probabilité sur $\mathbb{R}^d$, $\widehat{\mu * \nu} = \hat{\mu}\,\hat{\nu}$.)*
(iii) *Si X et Y sont de carré intégrable, $K_{X+Y} = K_X + K_Y$; En particulier, si $d=1$, $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$.*

Preuve. (i) Si X et Y sont indépendantes, on sait que $P_{(X,Y)} = P_X \otimes P_Y$, et donc, pour toute fonction mesurable positive φ sur $\mathbb{R}^d$,

$$E[\varphi(X+Y)] = \int \varphi(x+y)\, P_{(X,Y)}(dxdy) = \int\int \varphi(x+y)\, P_X(dx)P_Y(dy) = \int \varphi(z)\, P_X * P_Y(dz)$$

par définition de $P_X * P_Y$. Si de plus X et Y ont une densité,

$$E[\varphi(X+Y)] = \int\int \varphi(x+y)\, p_X(x)p_Y(y)dxdy = \int \varphi(z)\Big(\int p_X(x)p_Y(z-x)dx\Big)dz,$$

ce qui montre bien que $X+Y$ a pour densité $p_X * p_Y$ (remarquer que $p_X * p_Y$ est ici bien définie presque partout comme convolution de deux fonctions de $L^1(\mathbb{R}^d, \lambda)$).
(ii) Il suffit d'écrire

$$\Phi_{X+Y}(\xi) = E[e^{i\xi X}\, e^{i\xi Y}] = E[e^{i\xi X}]\, E[e^{i\xi Y}] = \Phi_X(\xi)\Phi_Y(\xi).$$

(iii) Si $X = (X_1, \ldots, X_d)$ et $Y = (Y_1, \ldots, Y_d)$, l'indépendance de X et Y entraîne que $\mathrm{cov}(X_i, Y_j) = 0$ pour tous $i, j \in \{1, \ldots, d\}$. En conséquence, par bilinéarité,

$$\mathrm{cov}(X_i + Y_i, X_j + Y_j) = \mathrm{cov}(X_i, X_j) + \mathrm{cov}(Y_i, Y_j)$$

ce qui donne bien $K_{X+Y} = K_X + K_Y$. □

Théorème 9.4.2 (Loi faible des grands nombres) *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi. Si $E[X_1^2] < \infty$, on a*

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n\to\infty]{L^2} E[X_1].$$

Preuve. Par linéarité,

$$E\Big[\frac{1}{n}(X_1 + \cdots + X_n)\Big] = E[X_1].$$

En conséquence,

$$E\Big[\Big(\frac{1}{n}(X_1 + \cdots + X_n) - E[X_1]\Big)^2\Big] = \frac{1}{n^2}\mathrm{var}(X_1 + \cdots + X_n) = \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{var}(X_j) = \frac{1}{n}\mathrm{var}(X_1)$$

qui tend vers 0 quand $n \to \infty$. □

Remarque. La preuve montre que le résultat reste vrai sous des hypothèses bien plus faibles. Au lieu de supposer que les v.a. X_n ont même loi, il suffit de demander que $E[X_n] = E[X_1]$ pour tout n et que la suite $E[X_n^2]$ soit bornée. Au lieu de l'indépendance, il suffit qu'on ait $\mathrm{cov}(X_n, X_m) = 0$ dès que $n \neq m$, ce qui est beaucoup plus faible.

Le mot "faible" dans la loi faible des grands nombres renvoie au fait que la convergence du théorème a lieu dans L^2, alors que d'un point de vue probabiliste il est plus significatif d'avoir une convergence presque sûre, c'est-à-dire une convergence simple en dehors d'un ensemble de probabilité nulle (on parle alors de loi forte). Nous donnons un premier énoncé allant dans ce sens, qui sera considérablement amélioré dans le chapitre suivant.

Proposition 9.4.3 *Reprenons les hypothèses du théorème précédent, et supposons de plus que $E[X_1^4] < \infty$. Alors on a presque sûrement*

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n\to\infty]{} E[X_1].$$

Preuve. Quitte à remplacer X_n par $X_n - E[X_n]$, on peut supposer que $E[X_n] = 0$. Alors,

$$E[(\frac{1}{n}(X_1 + \cdots + X_n))^4] = \frac{1}{n^4} \sum_{i_1,\ldots,i_4 \in \{1,\ldots,n\}} E[X_{i_1} X_{i_2} X_{i_3} X_{i_4}].$$

En utilisantl'indépendance et la propriété $E[X_k] = 0$, on voit que les seuls termes non nuls de la somme sont ceux pour lesquels chaque valeur prise par une composante du quadruplet (i_1, i_2, i_3, i_4) apparaît au moins deux fois dans ce quadruplet. En utilisant le fait que les X_k ont même loi, on trouve

$$E[(\frac{1}{n}(X_1 + \cdots + X_n))^4] = \frac{1}{n^4}\Big(nE[X_1^4] + 3n(n-1)E[X_1^2 X_2^2]\Big) \leq \frac{C}{n^2}$$

pour une certaine constante $C < \infty$. Il en découle que

$$\sum_{n=1}^{\infty} E[(\frac{1}{n}(X_1 + \cdots + X_n))^4] < \infty.$$

En intervertissant somme et espérance, on obtient

$$E\Big[\sum_{n=1}^{\infty} (\frac{1}{n}(X_1 + \cdots + X_n))^4\Big] < \infty,$$

d'où

$$\sum_{n=1}^{\infty} (\frac{1}{n}(X_1 + \cdots + X_n))^4 < \infty \;, \quad \text{p.s.}$$

ce qui entraîne l'assertion de la proposition. □

Corollaire 9.4.4 *Si $(A_n)_{n\geq 1}$ est une suite d'événements indépendants de même probabilité, on a*

$$\frac{1}{n} \sum_{i=1}^{n} 1_{A_i} \xrightarrow[n\to\infty]{\text{p.s.}} P(A_1).$$

Ce corollaire fait le lien entre notre approche axiomatique moderne et la définition historique de la probabilité comme fréquence d'apparition d'un événement quand on répète un grand nombre de fois une expérience aléatoire.

Revenons à la deuxième application du lemme de Borel-Cantelli donnée ci-dessus, qui concernait le développement dyadique

$$\omega = \sum_{k=1}^{\infty} X_k(\omega)\, 2^{-k}$$

d'un réel $\omega \in [0, 1[$. Si $p \geq 1$ est fixé, on a vu que les v.a. $Y_1 = (X_1, \ldots, X_p)$, $Y_2 = (X_{p+1}, \ldots, X_{2p}), \ldots$ sont indépendantes et de même loi. On déduit alors du corollaire que, pour tous $i_1, \ldots, i_p \in \{0, 1\}$,

$$d\omega \text{ p.s.} \qquad \frac{1}{n}\,\text{Card}\{j \leq n : Y_j(\omega) = (i_1, \ldots, i_p)\} \xrightarrow[n\to\infty]{} \frac{1}{2^p}.$$

Pour chaque $\ell \in \{1, \dots, p\}$, le même argument appliqué aux v.a. $(X_\ell, X_{\ell+1}, \dots, X_{p+\ell-1})$, $(X_{p+\ell}, X_{p+\ell+1}, \dots, X_{2p+\ell-1}), \dots$ conduit à

$$d\omega \text{ p.s.} \qquad \frac{1}{n}\,\mathrm{Card}\{j \le n : X_{jp+\ell}(\omega) = i_1, \dots, X_{(j+1)p+\ell-1}(\omega) = i_p\} \underset{n\to\infty}{\longrightarrow} \frac{1}{2^p}.$$

En combinant ces résultats on trouve

$$d\omega \text{ p.s.} \qquad \frac{1}{n}\,\mathrm{Card}\{k \le n : X_{k+1}(\omega) = i_1, \dots, X_{k+p}(\omega) = i_p\} \underset{n\to\infty}{\longrightarrow} \frac{1}{2^p}.$$

Comme une réunion dénombrable d'ensembles de probabilité nulle est encore de probabilité nulle, on a aussi, pour tout $\omega \in [0,1[$ sauf sur un ensemble de mesure nulle :

$$\forall p \ge 1, \forall i_1, \dots, i_p \in \{0,1\},\ \frac{1}{n}\mathrm{Card}\{k \le n : X_{k+1}(\omega) = i_1, \dots, X_{k+p}(\omega) = i_p\} \underset{n\to\infty}{\longrightarrow} \frac{1}{2^p}. \tag{9.3}$$

Autrement dit, pour presque tout réel ω de $[0,1[$, la fréquence d'apparition de n'importe quel bloc de longueur finie de 0 et de 1 dans le développement dyadique de ω existe et est égale à 2^{-p} si p est la longueur du bloc. Remarquons qu'il n'est pas facile d'exhiber un réel ω pour lequel la propriété (9.3) soit vraie. En fait, le moyen le plus rapide pour prouver que de tels réels existent est très certainement le raisonnement qui précède. Ceci est typique de l'application des probabilités à des problèmes d'existence : pour établir l'existence d'un objet ayant certaines propriétés, on montre qu'un objet pris au hasard (selon une loi de probabilité bien choisie) vérifie les propriétés en question.

Semigroupes de convolution

Soit $I = \mathbb{N}$ ou $I = \mathbb{R}_+$.

Définition 9.4.1 *Soit $(\mu_t)_{t\in I}$ une famille de mesures de probabilité sur $\mathbb{R}$ (ou sur $\mathbb{R}^d$). On dit que $(\mu_t)_{t\in I}$ est un semigroupe de convolution si $\mu_0 = \delta_0$ et si*

$$\mu_t * \mu_{t'} = \mu_{t+t'} \ , \quad \forall t, t' \in I.$$

L'interprétation probabiliste est que si X a pour loi μ_t, Y a pour loi $\mu_{t'}$ et si X et Y sont indépendantes, alors $X + Y$ a pour loi $\mu_{t+t'}$ (cf la première proposition de cette partie).

Lemme 9.4.5 *Pour que $(\mu_t)_{t\in I}$ soit un semigroupe de convolution, il suffit qu'il existe une fonction $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$ telle que :*

- *si $I = \mathbb{N}$, $\hat{\mu}_t(\xi) = \varphi(\xi)^t$, $\forall t \in I$;*
- *si $I = \mathbb{R}$, $\hat{\mu}_t(\xi) = \exp(-t\varphi(\xi))$, $\forall t \in I$.*

La preuve est immédiate puisque si $\hat{\mu}_t$ a la forme donnée, on a immédiatement

$$\hat{\mu}_{t+t'} = \hat{\mu}_t\, \hat{\mu}_{t'} = \widehat{\mu_t * \mu_{t'}}$$

et l'injectivité de la transformée de Fourier donne $\mu_{t+t'} = \mu_t * \mu_{t'}$.

Exemples.

(1) $I = \mathbb{N}$ et, pour tout $n \in \mathbb{N}^*$, μ_n est la loi binômiale $\mathcal{B}(n,p)$ (on a fixé $p \in [0,1]$). La propriété $\mu_{n+m} = \mu_n * \mu_m$ est immédiate à partir de l'interprétation probabiliste de la loi binômiale. Alternativement on peut utiliser le lemme en remarquant que $\hat{\mu}_n(\xi) = (pe^{i\xi} + 1 - p)^n$.

(2) $I = \mathbb{R}_+$ et, pour tout $t \in \mathbb{R}_+$, μ_t est la loi de Poisson de paramètre t. Dans ce cas,

$$\hat{\mu}_t(\xi) = \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{ik\xi} e^{-t} = \exp(-t(1 - e^{i\xi})).$$

(3) $I = \mathbb{R}_+$ et, pour tout $t > 0$, μ_t est la loi Gaussienne $\mathcal{N}(0,t)$. On a déjà calculé dans le Chapitre 8

$$\hat{\mu}_t(\xi) = \exp(-\frac{t\xi^2}{2}).$$

Conséquence importante. Si X et Y sont deux v.a. réelles indépendantes et

- si X suit la loi de Poisson de paramètre λ et X' la loi de Poisson de paramètre λ', alors $X + X'$ suit la loi de Poisson de paramètre $\lambda + \lambda'$;

- si X suit la loi gaussienne $\mathcal{N}(m, \sigma^2)$ et X' suit la loi gaussienne $\mathcal{N}(m', \sigma'^2)$, alors $X + X'$ suit la loi gaussienne $\mathcal{N}(m + m', \sigma^2 + \sigma'^2)$. (On se ramène au cas $m = m' = 0$ en considérant $X - m$ et $X' - m'$.)

Plus généralement toute combinaison linéaire de variables aléatoires gaussiennes indépendantes est encore gaussienne.

Chapitre 10

Convergence de variables aléatoires

La première partie de ce chapitre présente les différentes notions de convergence de variables aléatoires, et les liens existant entre ces notions. On établit ensuite la loi forte des grands nombres, qui est l'un des deux théorèmes limites fondamentaux de la théorie des probabilités. Le troisième paragraphe présente la convergence en loi des variables aléatoires : ce type de convergence est sans doute le plus délicat à comprendre, en partie parce qu'il s'agit d'une convergence de mesures (ce sont les lois des variables aléatoires qui convergent et non les variables elle-mêmes). La notion de convergence en loi, et le théorème important reliant cette convergence à celle des fonctions caractéristiques, permettent d'arriver au deuxième théorème limite fondamental qui est le théorème central limite.

10.1 Les différentes notions de convergence

Soient $(X_n)_{n\geq 1}$, X des variables aléatoires à valeurs dans $\mathbb{R}^d$, définies sur un espace de probabilité $(\Omega, \mathcal{A}, P)$. On a déjà rencontré plusieurs notions de convergence de la suite (X_n) vers X. En particulier

$$X_n \xrightarrow[n\to\infty]{\text{p.s.}} X \qquad \text{si } P(\{\omega \in \Omega : X(\omega) = \lim_{n\to\infty} X_n(\omega)\}) = 1,$$

et, pour $p \in [1, \infty[$,

$$X_n \xrightarrow[n\to\infty]{L^p} X \qquad \text{si } \lim_{n\to\infty} E[|X_n - X|^p] = 0.$$

Définition 10.1.1 *On dit que la suite (X_n) converge en probabilité vers X, et on note*

$$X_n \xrightarrow[n\to\infty]{\text{(P)}} X$$

si pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

Proposition 10.1.1 *Soit $\mathcal{L}^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, P)$ l'espace de toutes les variables aléatoires à valeurs dans $\mathbb{R}^d$, et soit $L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, P)$ son quotient par la relation d'équivalence $X \sim Y$ ssi $X = Y$ p.s. Alors, la formule*

$$d(X, Y) = E[|X - Y| \wedge 1]$$

définit une distance sur $L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, P)$ qui est compatible avec la convergence en probabilité, au sens où une suite (X_n) converge en probabilité vers X ssi $d(X_n, X)$ tend vers 0. De plus, l'espace $L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, P)$ est complet pour la distance d.

Preuve. Il est facile de vérifier que d est une distance. De plus, si la suite (X_n) converge en probabilité vers X, on a pour tout $\varepsilon > 0$,

$$E[|X_n - X| \wedge 1] \leq E[|X_n - X| 1_{\{|X_n - X| \leq \varepsilon\}}] + E[(|X_n - X| \wedge 1) 1_{\{|X_n - X| > \varepsilon\}}] \leq \varepsilon + P(|X_n - X| > \varepsilon).$$

D'après la définition de la convergence en probabilité, cela entraîne $\limsup d(X_n, X) \leq \varepsilon$, et puisque ε était arbitraire on a $d(X_n, X) \longrightarrow 0$. Inversement, si $d(X_n, X) \longrightarrow 0$, alors, pour tout $\varepsilon \in]0, 1]$,

$$P(|X_n - X| > \varepsilon) \leq \varepsilon^{-1} E[|X_n - X| \wedge 1] = \varepsilon^{-1} d(X_n, X) \underset{n\to\infty}{\longrightarrow} 0.$$

Il reste à voir que L^0 est complet pour la distance d. Soit donc (X_n) une suite de Cauchy pour la distance d. On peut trouver une sous-suite $Y_k = X_{n_k}$ telle que, pour tout $k \geq 1$,

$$d(Y_k, Y_{k+1}) \leq 2^{-k}.$$

Alors

$$E[\sum_{k=1}^{\infty} (|Y_{k+1} - Y_k| \wedge 1)] = \sum_{k=1}^{\infty} d(Y_k, Y_{k+1}) < \infty,$$

ce qui entraîne $\sum_{k=1}^{\infty} (|Y_{k+1} - Y_k| \wedge 1) < \infty$ p.s., et donc aussi $\sum_{k=1}^{\infty} |Y_{k+1} - Y_k| < \infty$ p.s. (p.s. il ne peut y avoir qu'un nombre fini de valeurs de k pour lesquelles $|Y_{k+1} - Y_k| \geq 1$). On définit ensuite une v.a. X dans L^0 en posant

$$X = Y_1 + \sum_{k=1}^{\infty} (Y_{k+1} - Y_k).$$

Par construction, la suite (Y_k) converge p.s. vers X, et cela entraîne

$$d(Y_k, X) = E[|Y_k - X| \wedge 1] \underset{k\to\infty}{\longrightarrow} 0,$$

par convergence dominée. Donc la suite (Y_k) converge en probabilité vers X, et cela est aussi vrai pour la suite de départ (X_n). □

La preuve précédente montre en particulier que de toute suite qui converge en probabilité on peut extraire une sous-suite qui converge p.s. (vers la même limite). Nous reprenons cette propriété dans l'énoncé suivant.

Proposition 10.1.2 *Si la suite (X_n) converge p.s., ou dans L^p, vers X, elle converge aussi en probabilité vers X. Inversement, si la suite (X_n) converge en probabilité vers X, il existe une sous-suite (X_{n_k}) qui converge p.s. vers X.*

Preuve. La deuxième assertion a déjà été vue. Pour la première, si X_n converge p.s. vers X,

$$d(X_n, X) = E[|X_n - X| \wedge 1] \underset{n\to\infty}{\longrightarrow} 0,$$

par convergence dominée. Si X_n converge dans L^p vers X,

$$d(X_n, X) \leq \|X_n - X\|_1 \leq \|X_n - X\|_p \underset{n\to\infty}{\longrightarrow} 0.$$

□

En résumé la convergence en probabilité est plus faible à la fois que la convergence p.s. et que la convergence dans L^p pour n'importe quel $p \in [1, \infty[$ (et a fortiori pour $p = \infty$). Dans l'autre sens, la convergence en probabilité entraîne la convergence p.s. pour une sous-suite, et la proposition ci-dessous donne des conditions qui permettent de déduire la convergence L^p de la convergence en probabilité.

Proposition 10.1.3 *Soit (X_n) une suite de v.a. convergeant en probabilité vers X. Supposons qu'il existe $r \in]1, \infty[$ tel que la suite (X_n) soit bornée dans L^r. Alors, pour tout $p \in [1, r[$, la suite (X_n) converge vers X dans L^p.*

Preuve. Par hypothèse, il existe une constante C telle que $E[|X_n|^r] \leq C$ pour tout n. Le lemme de Fatou entraîne alors $E[|X|^r] \leq C$ et donc $X \in L^r$. Ensuite, en utilisant l'inégalité de Hölder, on a pour tout $p \in [1, r[$ et tout $\varepsilon > 0$,

$$\begin{aligned} E[|X_n - X|^p] &= E[|X_n - X|^p 1_{\{|X_n - X| \leq \varepsilon\}}] + E[|X_n - X|^p 1_{\{|X_n - X| > \varepsilon\}}] \\ &\leq \varepsilon^p + E[|X_n - X|^r]^{p/r}\, P(|X_n - X| > \varepsilon)^{1-p/r} \\ &\leq \varepsilon^p + 2^p C^{p/r}\, P(|X_n - X| > \varepsilon)^{1-p/r}. \end{aligned}$$

En utilisant l'hypothèse de convergence en probabilité, il vient

$$\limsup_{n\to\infty} E[|X_n - X|^p] \leq \varepsilon^p$$

d'où le résultat annoncé puisque ε est arbitraire. □

10.2 La loi forte des grands nombres

Notre objectif est de montrer que si (X_n) est une suite de v.a. indépendantes et de même loi, dans L^1, alors les moyennes $\frac{1}{n}(X_1 + \cdots + X_n)$ convergent p.s. vers $E[X_1]$. Nous avons déjà obtenu ce résultat sous l'hypothèse supplémentaire que $E[|X_1|^4] < \infty$, mais nous cherchons maintenant à l'établir sous des hypothèses optimales. Nous commençons par un résultat préliminaire important.

Théorème 10.2.1 (Loi du tout ou rien) *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes, à valeurs dans des espaces mesurables quelconques. Pour tout $n \geq 1$ soit $\mathcal{B}_n$ la tribu*

$$\mathcal{B}_n = \sigma(X_k\, ;\, k \geq n).$$

Alors la tribu asymptotique $\mathcal{B}_\infty$ définie par

$$\mathcal{B}_\infty = \bigcap_{n=1}^{\infty} \mathcal{B}_n$$

est grossière, au sens où $P(B) = 0$ ou 1 pour tout $B \in \mathcal{B}_\infty$.

Preuve. Posons

$$\mathcal{D}_n = \sigma(X_k\,;\, k \le n).$$

On a observé dans le Chapitre 9 que pour tout n, $\mathcal{D}_n$ est indépendante de $\mathcal{B}_{n+1}$, donc a fortiori de $\mathcal{B}_\infty$. Ainsi,

$$\forall A \in \bigcup_{n=1}^{\infty} \mathcal{D}_n\ ,\ \forall B \in \mathcal{B}_\infty\ , \quad P(A \cap B) = P(A)P(B).$$

Puisque la classe $\bigcup_{n=1}^{\infty} \mathcal{D}_n$ est stable par intersections finies, un autre résultat du Chapitre 9 permet alors de conclure que $\mathcal{B}_\infty$ est indépendante de

$$\sigma\Big(\bigcup_{n=1}^{\infty} \mathcal{D}_n\Big) = \sigma(X_n\,;\, n \ge 1).$$

En particulier, $\mathcal{B}_\infty$ est indépendante d'elle-même, et pour tout $B \in \mathcal{B}_\infty$, $P(B) = P(B \cap B) = P(B)^2$, ce qui n'est possible que si $P(B) = 0$ ou 1. □

On vérifie aisément qu'une v.a. réelle mesurable par rapport à une tribu grossière est constante p.s. (sa fonction de répartition ne peut prendre que les deux valeurs 0 ou 1). On peut appliquer le théorème précédent à toute suite $(X_n)_{n\ge1}$ de v.a. réelles indépendantes. Il est facile de voir que la v.a.

$$\limsup_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n)$$

est mesurable par rapport à $\mathcal{B}_\infty$, et cela entraîne que cette variable (à valeurs dans $[-\infty, \infty]$) est constante p.s. En particulier, si on sait que la suite $\frac{1}{n}(X_1 + \cdots + X_n)$ converge p.s. la limite est constante (p.s.).

Avant d'utiliser la loi du tout ou rien pour établir la loi forte des grands nombres, nous donnons d'abord une application plus facile au jeu de pile ou face.

Proposition 10.2.2 *Soit $(X_n)_{n\ge1}$ une suite de variables aléatoires indépendantes , de même loi donnée par $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$. Pour tout $n \ge 1$, posons*

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Alors,

$$p.s. \qquad \sup_{n\ge1} S_n = +\infty \quad et \quad \inf_{n\ge1} S_n = -\infty.$$

En particulier, il existe p.s. des entiers n arbitrairement grands tels que $S_n = 0$.

En d'autres termes si on imagine un jeu où à chaque instant entier le joueur gagne ou perd un Euro avec probabilité $1/2$, S_n représente le gain (positif ou négatif) accumulé après n instants. La proposition montre que quand $n \to \infty$, S_n prend tantôt des valeurs positives tantôt des valeurs négatives, de plus en plus grandes en valeur absolue.

Preuve. On commence par montrer que, pour tout entier $p \geq 1$,

$$P(-p \leq \inf_n S_n \leq \sup_n S_n \leq p) = 0.$$

Pour cela on fixe un entier $k > 2p$, et on remarque que

$$\bigcup_{j=0}^{\infty} \{X_{jk+1} = X_{jk+2} = \cdots = X_{jk+k} = 1\} \subset (\{-p \leq \inf_n S_n \leq \sup_n S_n \leq p\})^c.$$

Or une application du lemme de Borel-Cantelli (cf le Chapitre 9 pour des raisonnements analogues) montre que l'ensemble de gauche a probabilité 1, ce qui donne le résultat annoncé.

En faisant tendre p vers ∞, on trouve

$$P(\{\inf_n S_n > -\infty\} \cap \{\sup_n S_n < \infty\}) = 0,$$

d'où

$$P(\{\inf_n S_n = -\infty\} \cup \{\sup_n S_n = \infty\}) = 1,$$

et en particulier

$$P(\{\inf_n S_n = -\infty\}) + P(\{\sup_n S_n = \infty\}) \geq 1.$$

Un argument de symétrie montre que

$$P(\{\inf_n S_n = -\infty\}) = P(\{\sup_n S_n = \infty\})$$

et d'après ce qui précède ces deux probabilités sont strictement positives. Pour conclure, on remarque que

$$\{\sup_n S_n = \infty\} \in \mathcal{B}_\infty.$$

En effet, pour tout entier $k \geq 1$,

$$\{\sup_n S_n = \infty\} = \{\sup_{n \geq k}(X_k + X_{k+1} + \cdots + X_n) = \infty\} \in \mathcal{B}_k$$

et donc l'événement $\{\sup_n S_n = \infty\}$ est mesurable par rapport à l'intersection des tribus $\mathcal{B}_k$, c'est-à-dire $\mathcal{B}_\infty$. La loi du tout ou rien montre alors que $P(\{\sup_n S_n = \infty\}) = 1$. □

Nous passons maintenant au résultat principal de ce paragraphe.

Théorème 10.2.3 (Loi forte des grands nombres) *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, de même loi, dans L^1. Alors,*

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n \to \infty]{\text{p.s.}} E[X_1].$$

Remarques. (i) L'hypothèse d'intégrabilité est optimale dans le sens où elle est nécessaire pour que la limite $E[X_1]$ soit bien définie (et finie). Dans le cas où les v.a. X_n sont positives et $E[X_1] = \infty$, on montre facilement que

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n\to\infty]{\text{p.s.}} +\infty$$

en appliquant le théorème aux v.a. $X_n \wedge K$.

(ii) On peut montrer que la convergence du théorème a aussi lieu dans L^1. Nous ne donnerons pas la preuve ici (elle sera donnée à la fin du chapitre 12 en application de la théorie des martingales). Du point de vue probabiliste, c'est la convergence presque sûre qui a le plus de signification.

Preuve. Pour alléger les notations on pose $S_n = X_1 + \cdots + X_n$, $S_0 = 0$. Soit $a > E[X_1]$, et

$$M = \sup_{n\in\mathbb{N}} (S_n - na)$$

qui est une v.a. à valeurs dans $[0, \infty]$. Nous allons montrer que

$$M < \infty \ , \quad \text{p.s.} \tag{10.1}$$

Puisque l'inégalité $S_n \leq na + M$ est vraie pour tout n, il en découle aussitôt que

$$\limsup_{n\to\infty} \frac{1}{n} S_n \leq a \ , \quad \text{p.s.}$$

En considérant une suite de valeurs de a qui décroît vers $E[X_1]$, on trouve alors

$$\limsup_{n\to\infty} \frac{1}{n} S_n \leq E[X_1] \ , \quad \text{p.s.}$$

En remplaçant X_n par $-X_n$, on obtient l'inégalité inverse

$$\liminf_{n\to\infty} \frac{1}{n} S_n \geq E[X_1] \ , \quad \text{p.s.}$$

et l'énoncé du théorème découle de ces deux dernières inégalités.

Il reste à montrer (10.1). On remarque d'abord que, avec les notations de la loi du tout ou rien, l'événement $\{M < \infty\}$ est dans la tribu $\mathcal{B}_\infty$. En effet, il suffit d'écrire pour tout entier $k \geq 0$,

$$\{M < \infty\} = \{\sup_{n\in\mathbb{N}} (S_n - na) < \infty\} = \{\sup_{n\geq k} (S_n - S_k - (n-k)a) < \infty\}$$

et de remarquer que le dernier événement est mesurable pour la tribu $\sigma(X_{k+1}, X_{k+2}, \ldots)$. Pour conclure il suffira donc de montrer que $P(M < \infty) > 0$, ou de manière équivalente que $P(M = \infty) < 1$, ce que nous ferons en raisonnant par l'absurde.

Commençons par quelques notations. Pour tout entier $k \in \mathbb{N}$, posons

$$M_k = \sup_{0 \le n \le k} (S_n - na),$$
$$M'_k = \sup_{0 \le n \le k} (S_{n+1} - S_1 - na).$$

Alors M_k et M'_k ont même loi : en effet d'une part les vecteurs $(X_1, \dots, X_k)$ et $(X_2, \dots, X_{k+1})$ ont même loi et d'autre part on peut écrire $M_k = F_k(X_1, \dots, X_k)$ et $M'_k = F_k(X_2, \dots, X_{k+1})$ avec la même fonction (déterministe) $F_k : \mathbb{R}^k \longrightarrow \mathbb{R}$. Il en découle que

$$M = \lim_{k \to \infty} \uparrow M_k$$

et

$$M' = \lim_{k \to \infty} \uparrow M'_k$$

ont aussi même loi (écrire $P(M' \le x) = \lim \downarrow P(M'_k \le x) = \lim \downarrow P(M_k \le x) = P(M \le x)$).

Par ailleurs, il découle des définitions que pour tout $k \ge 1$,

$$M_{k+1} = \sup\Big(0, \sup_{1 \le n \le k+1} (S_n - na)\Big) = \sup(0, M'_k + X_1 - a),$$

ce qu'on peut encore réécrire sous la forme

$$M_{k+1} = M'_k - \inf(a - X_1, M'_k).$$

Puisque M'_k a même loi que M_k (et que ces deux v.a. sont clairement dans L^1), on trouve

$$E[\inf(a - X_1, M'_k)] = E[M'_k] - E[M_{k+1}] = E[M_k] - E[M_{k+1}] \le 0$$

grâce à l'inégalité triviale $M_k \le M_{k+1}$. On peut maintenant appliquer le théorème de convergence dominée à la suite des v.a. $\inf(a - X_1, M'_k)$, qui sont dominées en valeur absolue par $|a - X_1|$ (rappelons que $M'_k \ge 0$). Il vient alors

$$E[\inf(a - X_1, M')] = \lim_{k \to \infty} E[\inf(a - X_1, M'_k)] \le 0.$$

Si on avait $P(M = \infty) = 1$, on aurait aussi $P(M' = \infty) = 1$, puisque les v.a. M et M' ont même loi, et donc $\inf(a - X_1, M') = a - X_1$ p.s. Mais alors l'inégalité précédente donnerait $E[a - X_1] \le 0$, ce qui est absurde puisqu'on a choisi $a > E[X_1]$. Cette contradiction termine la preuve. □

10.3 La convergence en loi

Rappelons que $C_b(\mathbb{R}^d)$ désigne l'espace des fonctions continues bornées de $\mathbb{R}^d$ dans $\mathbb{R}$, qu'on munit de la norme sup

$$\|\varphi\| = \sup_{x \in \mathbb{R}^d} |\varphi(x)|.$$

Définition 10.3.1 *Une suite* (μ_n) *de mesures de probabilité sur* $\mathbb{R}^d$ *converge étroitement vers une mesure de probabilité* μ *sur* $\mathbb{R}^d$ *(on note* $\mu_n \xrightarrow{(e)} \mu$*) si*

$$\forall \varphi \in C_b(\mathbb{R}^d) \ , \qquad \int \varphi \, d\mu_n \underset{n\to\infty}{\longrightarrow} \int \varphi \, d\mu.$$

Une suite (X_n) *de v.a. à valeurs dans* $\mathbb{R}^d$ *converge en loi vers une v.a.* X *à valeurs dans* $\mathbb{R}^d$ *(on note* $X_n \xrightarrow{(\text{loi})} X$*) si la suite* (P_{X_n}) *converge étroitement vers* P_X*. Cela équivaut encore à*

$$\forall \varphi \in C_b(\mathbb{R}^d) \ , \qquad E[\varphi(X_n)] \underset{n\to\infty}{\longrightarrow} E[\varphi(X)].$$

Remarques. (i) Il y a un abus de langage à dire que la suite de v.a. (X_n) converge en loi vers X, car la v.a. limite X n'est pas définie de manière unique : seule sa loi P_X l'est (pour cette raison on écrira parfois qu'une suite de v.a. (X_n) converge en loi vers μ mesure de probabilité sur $\mathbb{R}^d$, et il faudra évidemment comprendre que la suite (P_{X_n}) converge étroitement vers μ). Notons aussi qu'on peut considérer la convergence en loi de v.a. définies sur des espaces de probabilité différents (ici nous supposerons toujours implicitement qu'elles sont définies sur le même espace de probabilité), ce qui rend la convergence en loi très différente des autres convergences discutées ci-dessus.

(ii) L'espace des mesures de probabilité sur $\mathbb{R}^d$ peut être vu comme un sous-ensemble du dual $C_b(\mathbb{R}^d)^*$. La convergence étroite correspond alors à la topologie faible * sur le dual (topologie de la convergence simple, les éléments du dual étant vus comme des fonctions sur $C_b(\mathbb{R}^d)$).

Exemples. (a) Si les v.a. X_n et X sont à valeurs dans $\mathbb{Z}^d$, alors X_n converge en loi vers X si et seulement si

$$\forall x \in \mathbb{Z}^d \ , \quad P(X_n = x) \underset{n\to\infty}{\longrightarrow} P(X = x)$$

(l'implication $\Leftarrow$ demande un petit raisonnement : l'argument est facile si on sait, ce qui sera établi plus tard, qu'on peut remplacer $C_b(\mathbb{R}^d)$ par $C_c(\mathbb{R}^d)$ dans la définition de la convergence étroite).

(b) Si les X_n sont des v.a. à densité, $P_{X_n}(dx) = p_n(x)dx$, si on suppose

$$p_n(x) \longrightarrow p(x) \ , \quad dx \text{ p.p.}$$

et s'il existe une fonction $q \geq 0$ telle que $\int_{\mathbb{R}^d} q(x)dx < \infty$ et

$$\forall n \ , \quad p_n(x) \leq q(x) \ , \quad dx \text{ p.p.}$$

alors p est une densité de probabilité sur $\mathbb{R}^d$, et X_n converge en loi vers la loi $p(x)dx$. Cela découle du théorème de convergence dominée.

(c) Si X_n est de loi uniforme sur $\{\frac{1}{2^n}, \frac{2}{2^n}, \ldots, \frac{2^n}{2^n}\}$, alors X_n converge en loi vers la loi uniforme sur $[0,1]$. Ce résultat découle de l'approximation de l'intégrale d'une fonction continue par ses sommes de Riemann.

(d) Si X_n est de loi gaussienne $\mathcal{N}(0, \sigma_n^2)$ et si $\sigma_n \longrightarrow 0$, alors X_n converge en loi vers la v.a. constante égale à 0.

Proposition 10.3.1 *Si la suite (X_n) converge en probabilité vers X alors la suite (X_n) converge en loi vers X.*

Preuve. Supposons d'abord que X_n converge p.s. vers X. Alors, pour toute fonction $\varphi \in C_b(\mathbb{R}^d)$, $\varphi(X_n)$ converge p.s. vers $\varphi(X)$ et donc le théorème de convergence dominée entraîne $E[\varphi(X_n)] \longrightarrow E[\varphi(X)]$, d'où la convergence en loi recherchée.

Dans le cas général, raisonnons par l'absurde en supposant que X_n ne converge pas en loi vers X, donc qu'il existe une fonction $\varphi \in C_b(\mathbb{R}^d)$ telle que $E[\varphi(X_n)]$ ne converge pas vers $E[\varphi(X)]$. On peut trouver une sous-suite (n_k) et $\varepsilon > 0$ tels que $|E[\varphi(X_{n_k})] - E[\varphi(X)]| \geq \varepsilon$ pour tout k. Mais, d'après un résultat de la partie 1, il existe une sous-sous-suite (n_{k_ℓ}) telle que $(X_{n_{k_\ell}})$ converge p.s. vers X. La première partie de la preuve donne alors une contradiction. □

Remarque. Il existe un cas où la réciproque de la proposition est vraie. C'est le cas où la v.a. limite X est constante (p.s.). En effet, si X_n converge en loi vers $a \in \mathbb{R}^d$, il découle de la propriété (ii) de la proposition qui suit que pour tout $\varepsilon > 0$,

$$\liminf_{n\to\infty} P_{X_n}(B(a,\varepsilon)) \geq 1$$

où $B(a,\varepsilon)$ est la boule ouverte de centre a et de rayon ε. C'est exactement dire que X_n converge en probabilité vers a.

Si (X_n) est une suite de v.a. convergeant en loi vers X, il n'est pas toujours vrai qu'on ait

$$P(X_n \in B) \longrightarrow P(X \in B)$$

pour tout borélien B de $\mathbb{R}^d$ (prendre $B = \{0\}$ dans l'exemple (d) ci-dessus). On a cependant le résultat suivant.

Proposition 10.3.2 *Soient $(\mu_n), \mu$ des mesures de probabilité sur $\mathbb{R}^d$. Les quatre assertions suivantes sont équivalentes.*

(i) *La suite (μ_n) converge étroitement vers μ.*

(ii) *Pour tout ouvert G de $\mathbb{R}^d$,*
$$\liminf \mu_n(G) \geq \mu(G).$$

(iii) *Pour tout fermé F de $\mathbb{R}^d$,*
$$\limsup \mu_n(F) \leq \mu(F).$$

(iv) *Pour tout borélien B de $\mathbb{R}^d$ tel que $\mu(\partial B) = 0$,*
$$\lim \mu_n(B) = \mu(B).$$

Preuve. Commençons par montrer (i)⇒(ii). Si G est un ouvert de $\mathbb{R}^d$, on peut trouver une suite (φ_p) de fonctions continues bornées telles que $0 \leq \varphi_p \leq 1_G$ et $\varphi_p \uparrow 1_G$ (par exemple $\varphi_p(x) = p\,\mathrm{dist}(x, G^c) \wedge 1$). Alors,

$$\liminf_{n\to\infty} \mu_n(G) \geq \sup_p \Big(\liminf_{n\to\infty} \int \varphi_p d\mu_n\Big) = \sup_p \Big(\int \varphi_p d\mu\Big) = \mu(G).$$

L'équivalence (ii)$\Leftrightarrow$(iii) est immédiate par passage au complémentaire.

Montrons que (ii) et (iii) entraînent (iv). Si $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\limsup \mu_n(B) \leq \limsup \mu_n(\overline{B}) \leq \mu(\overline{B})$$
$$\liminf \mu_n(B) \geq \liminf \mu_n(\overset{\circ}{B}) \geq \mu(\overset{\circ}{B}).$$

Si $\mu(\partial B) = 0$ on a $\mu(\overline{B}) = \mu(\overset{\circ}{B}) = \mu(B)$ et on obtient (iv).

Il reste à montrer l'implication (iv)$\Rightarrow$(i). Soit $\varphi \in C_b(\mathbb{R}^d)$. Quitte à décomposer $\varphi = \varphi^+ - \varphi^-$ on peut supposer $\varphi \geq 0$. Soit $K > 0$ tel que $0 \leq \varphi \leq K$. Alors le théorème de Fubini montre que

$$\int \varphi(x)\mu(dx) = \int \Big(\int_0^K 1_{\{t \leq \varphi(x)\}} dt \Big) \mu(dx) = \int_0^K \mu(E_t^\varphi) dt,$$

où $E_t^\varphi = \{x \in \mathbb{R}^d : \varphi(x) \geq t\}$. De même, pour tout n,

$$\int \varphi(x)\mu_n(dx) = \int_0^K \mu_n(E_t^\varphi) dt.$$

Remarquons que $\partial E_t^\varphi \subset \{x \in \mathbb{R}^d : \varphi(x) = t\}$, et qu'il existe au plus une infinité dénombrable de valeurs de t telles que

$$\mu(\{x \in \mathbb{R}^d : \varphi(x) = t\}) > 0$$

(en effet il y a au plus k valeurs distinctes de t telles que $\mu(\{x \in \mathbb{R}^d : \varphi(x) = t\}) \geq \frac{1}{k}$). Donc (iv) entraîne

$$\mu_n(E_t^\varphi) \underset{n\to\infty}{\longrightarrow} \mu(E_t^\varphi) \;, \quad dt \text{ p.p.}$$

et par convergence dominée on obtient

$$\int \varphi(x)\mu_n(dx) = \int_0^K \mu_n(E_t^\varphi) dt \underset{n\to\infty}{\longrightarrow} \int_0^K \mu_n(E_t^\varphi) dt = \int \varphi(x)\mu(dx).$$

□

Conséquence. Une suite (X_n) de v.a. réelles converge en loi vers une v.a. X si et seulement si les fonctions de répartition $F_{X_n}(x)$ convergent vers $F_X(x)$ en tout point x où F_X est continue. L'implication $\Rightarrow$ découle immédiatement de la propriété (iv) ci-dessus. Dans l'autre sens, on observe que sous la condition de convergence des fonctions de répartition (en tout point où F_X est continue), on a pour tout $x \in \mathbb{R}$,

$$\liminf F_{X_n}(x-) \geq F_X(x-),$$
$$\limsup F_{X_n}(x) \leq F_X(x).$$

Il découle de cette observation que la condition (ii) de la proposition est satisfaite pour $\mu_n = P_{X_n}$ et $\mu = P_X$ lorsque G est un intervalle ouvert. Il suffit ensuite d'écrire un ouvert quelconque comme réunion dénombrable disjointe d'intervalles ouverts pour aboutir au résultat désiré.

Rappelons la notation $C_c(\mathbb{R}^d)$ pour l'espace des fonctions continues à support compact sur $\mathbb{R}^d$.

Proposition 10.3.3 *Soient (μ_n) et μ des mesures de probabilité sur $\mathbb{R}^d$. Soit H un sous-ensemble de $C_b(\mathbb{R}^d)$ dont l'adhérence (pour la norme sup) contient $C_c(\mathbb{R}^d)$. Les propriétés suivantes sont équivalentes :*

(i) *La suite (μ_n) converge étroitement vers μ.*

(ii) *On a*
$$\forall \varphi \in C_c(\mathbb{R}^d) \ , \quad \int \varphi \, d\mu_n \underset{n\to\infty}{\longrightarrow} \int \varphi \, d\mu.$$

(iii) *On a*
$$\forall \varphi \in H \ , \quad \int \varphi \, d\mu_n \underset{n\to\infty}{\longrightarrow} \int \varphi \, d\mu.$$

Preuve. Il est évident que (i)⇒(ii) et (i)⇒(iii). Supposons ensuite que (ii) est satisfaite. Soit $\varphi \in C_b(\mathbb{R}^d)$ et soit (f_k) une suite de fonctions dans $C_c(\mathbb{R}^d)$ telles que $0 \leq f_k \leq 1$ et $f_k \uparrow 1$ quand $k \to \infty$. Alors pour tout k, $\varphi f_k \in C_c(\mathbb{R}^d)$ et donc
$$\int \varphi f_k \, d\mu_n \underset{n\to\infty}{\longrightarrow} \int \varphi f_k \, d\mu.$$

Par ailleurs,
$$\begin{aligned}
\Big| \int \varphi d\mu_n - \int \varphi f_k \, d\mu_n \Big| &\leq \Big(\sup_{x\in\mathbb{R}} |\varphi(x)| \Big) \Big(1 - \int f_k d\mu_n \Big), \\
\Big| \int \varphi d\mu - \int \varphi f_k \, d\mu \Big| &\leq \Big(\sup_{x\in\mathbb{R}} |\varphi(x)| \Big) \Big(1 - \int f_k d\mu \Big).
\end{aligned}$$

Donc, pour tout k,
$$\begin{aligned}
\limsup_{n\to\infty} \Big| \int \varphi \, d\mu_n - \int \varphi \, d\mu \Big| &\leq \Big(\sup_{x\in\mathbb{R}} |\varphi(x)| \Big) \Big(\limsup_{n\to\infty} (1 - \int f_k d\mu_n) + (1 - \int f_k d\mu) \Big) \\
&= 2 \Big(\sup_{x\in\mathbb{R}} |\varphi(x)| \Big) (1 - \int f_k d\mu) \Big).
\end{aligned}$$

Il suffit maintenant de faire tendre k vers ∞ pour trouver que $\int \varphi d\mu_n$ converge vers $\int \varphi d\mu$, et on a établi (i).

Il reste à montrer (iii)⇒(ii). On suppose donc que la propriété (iii) est satisfaite. Ensuite, si $\varphi \in C_c(\mathbb{R}^d)$, on peut pour chaque entier $k \geq 1$ trouver une fonction $\varphi_k \in H$ telle que $\|\varphi - \varphi_k\| \leq 1/k$. Mais alors, pour tout $k \geq 1$,
$$\begin{aligned}
&\limsup_{n\to\infty} | \int \varphi d\mu_n - \int \varphi d\mu | \\
&\leq \limsup_{n\to\infty} \Big(| \int \varphi d\mu_n - \int \varphi_k d\mu_n | + | \int \varphi_k d\mu_n - \int \varphi_k d\mu | + | \int \varphi_k d\mu - \int \varphi d\mu | \Big) \leq \frac{2}{k}.
\end{aligned}$$

Comme k est arbitraire cela donne $\int \varphi d\mu_n \longrightarrow \int \varphi d\mu$, d'où la propriété (ii). □

Théorème 10.3.4 (Lévy) *Une suite (μ_n) de mesures de probabilité sur $\mathbb{R}^d$ converge étroitement vers une mesure de probabilité μ sur $\mathbb{R}^d$ si et seulement si*

$$\forall \xi \in \mathbb{R}^d \;, \quad \widehat{\mu}_n(\xi) \underset{n\to\infty}{\longrightarrow} \widehat{\mu}(\xi).$$

De manière équivalente, une suite (X_n) de variables aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers X si et seulement si

$$\forall \xi \in \mathbb{R}^d \;, \quad \Phi_{X_n}(\xi) \underset{n\to\infty}{\longrightarrow} \Phi_X(\xi).$$

Preuve. Il suffit de montrer la première assertion. D'abord, si on suppose que la suite (μ_n) converge étroitement vers μ, la définition même de cette convergence assure que

$$\forall \xi \in \mathbb{R}^d \;, \quad \widehat{\mu}_n(\xi) = \int e^{i\xi\cdot x}\mu_n(dx) \underset{n\to\infty}{\longrightarrow} \int e^{i\xi\cdot x}\mu(dx) = \widehat{\mu}(\xi).$$

Supposons inversement que $\widehat{\mu}_n(\xi) \to \widehat{\mu}(\xi)$ pour tout $\xi \in \mathbb{R}^d$ et montrons qu'alors la suite (μ_n) converge étroitement vers μ. Pour alléger l'écriture on traite seulement le cas $d = 1$.

Soit $f \in C_c(\mathbb{R})$ et pour tout $\sigma > 0$ soit

$$g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}\, \exp(-\frac{x^2}{2\sigma^2}).$$

Alors on a déjà observé à la fin du Chapitre 8 que $g_\sigma * f$ converge simplement vers f quand $\sigma \to 0$. En fait on vérifie aisément que cette convergence est uniforme sur $\mathbb{R}$.

Par ailleurs, si ν est une mesure de probabilité sur $\mathbb{R}$, on a vu dans la preuve du théorème d'injectivité de la transformée de Fourier (fin du Chapitre 8) que

$$\int g_\sigma * f \, d\nu = \int f(x)\, g_\sigma * \nu(x) dx = \int f(x)\Big((\sigma\sqrt{2\pi})^{-1}\int e^{i\xi x} g_{1/\sigma}(\xi)\widehat{\nu}(-\xi)d\xi\Big)dx.$$

Puisque $\widehat{\mu}_n(\xi) \to \widehat{\mu}(\xi)$ pour tout $\xi \in \mathbb{R}$, le théorème de convergence dominée entraîne que

$$\int e^{i\xi x} g_{1/\sigma}(\xi)\widehat{\mu}_n(-\xi)d\xi \underset{n\to\infty}{\longrightarrow} \int e^{i\xi x} g_{1/\sigma}(\xi)\widehat{\mu}(-\xi)d\xi,$$

et puisque ces quantités sont bornées en module par 1, on peut utiliser la formule précédente et à nouveau le théorème de convergence dominée pour obtenir que

$$\int g_\sigma * f \, d\mu_n \underset{n\to\infty}{\longrightarrow} \int g_\sigma * f \, d\mu.$$

Finalement, soit H le sous-espace de $C_b(\mathbb{R}^d)$ défini par

$$H = \{\varphi = g_\sigma * f : \sigma > 0 \text{ et } f \in C_c(\mathbb{R}^d)\}.$$

Alors l'adhérence de H dans $C_b(\mathbb{R}^d)$ contient $C_c(\mathbb{R}^d)$ (on a remarqué que si $f \in C_c(\mathbb{R}^d)$, $\|g_\sigma * f - f\|$ tend vers 0 quand $\sigma \to 0$) et on vient de montrer que $\int \varphi d\mu_n \longrightarrow \int \varphi d\mu$ pour toute fonction $\varphi \in H$. D'après la proposition précédente, cela suffit pour donner la convergence étroite de la suite (μ_n) vers μ. □

10.4 Deux applications

10.4.1 La convergence des mesures empiriques

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$, indépendantes et de même loi. Ces variables peuvent représenter les résultats successifs d'une même expérience aléatoire répétée de manière indépendante. Un problème statistique fondamental est d'estimer la loi de X_1 à partir de la donnée de $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$ pour une *seule* valeur de ω.

Exemple : théorie des sondages. Imaginons qu'on a une population de N individus numérotés $1, 2, \ldots, N$. L'entier N est supposé "très grand" (on peut penser à la population française). A l'individu i est attaché un paramètre $a(i) \in \mathbb{R}^d$ (par exemple, l'âge de l'individu, son intention de vote, son revenu mensuel, etc.). Si $A \in \mathcal{B}(\mathbb{R}^d)$, on s'intéresse alors à la quantité

$$\mu(A) = \frac{1}{N} \sum_{i=1}^{N} 1_A(a(i))$$

qui est la proportion d'individus dans la population dont le paramètre est dans A (par exemple la proportion d'individus de plus de cinquante ans qui ont l'intention de voter Chirac et ont un revenu mensuel supérieur à 2000 Euros).

Comme N est très grand, il est hors de question de calculer exactement $\mu(A)$. Le principe d'un sondage est alors de choisir un échantillon de la population, c'est-à-dire de prendre au hasard n individus (n grand mais petit devant N) en espérant que la proportion d'individus choisis dans cet échantillon pour lesquels le paramètre est dans A sera proche de la même proportion calculée pour la population totale. Pour rendre ceci précis en termes mathématiques, on se donne une famille $Y_1, \ldots, Y_n$ de variables aléatoires indépendantes de loi uniforme sur $\{1, \ldots, N\}$ (ce sont les individus de notre échantillon). La valeur du paramètre pour l'individu Y_j est $X_j = a(Y_j)$. Les v.a. $X_1, \ldots, X_n$ sont évidemment indépendantes et de même loi. De plus, cette loi est

$$P_{X_1}(A) = P(a(Y_1) \in A) = \frac{1}{N} \sum_{i=1}^{N} 1_A(a(i)) = \mu(A).$$

Par ailleurs, la proportion calculée sur les individus de l'échantillon est

$$\frac{1}{n} \sum_{j=1}^{n} 1_A(X_j(\omega)) = \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j(\omega)}(A)$$

Finalement, la question de savoir si la proportion calculée sur l'échantillon est proche de la proportion réelle $\mu(A)$ se ramène à vérifier que la mesure, dite "mesure empirique",

$$\frac{1}{n} \sum_{j=1}^{n} \delta_{X_j(\omega)}$$

est proche de P_{X_1} quand $n \to \infty$. Le théorème suivant apporte une réponse à cette question.

Théorème 10.4.1 *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et de même loi, à valeurs dans $\mathbb{R}^d$. Pour tout $\omega \in \Omega$ et tout $n \geq 1$, soit $\mu_{n,\omega}$ la mesure de probabilité sur $\mathbb{R}^d$ définie par*

$$\mu_{n,\omega} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(\omega)}.$$

Alors, p.s.,

$$\mu_{n,\omega} \xrightarrow[n\to\infty]{\text{(e)}} P_{X_1}.$$

Remarque. D'un point de vue pratique, le théorème précédent n'a aucun intérêt si on n'a pas d'estimation de la vitesse de convergence. En revenant à l'exemple donné avant l'énoncé du théorème, il faut que la mesure empirique $\mu_{n,\omega}$ soit "suffisamment proche" de P_{X_1} pour des valeurs de n grandes mais petites devant la taille N de la population (en pratique, N est de l'ordre de 10^7 et n seulement de l'ordre de 10^3).

Preuve. Soit H un sous-ensemble dénombrable dense de $C_c(\mathbb{R}^d)$. Si $\varphi \in H$, la loi forte des grands nombres appliquée aux v.a. $\varphi(X_i)$ assure que

$$\frac{1}{n} \sum_{i=1}^{n} \varphi(X_i) \xrightarrow[n\to\infty]{\text{p.s.}} E[\varphi(X_1)].$$

On peut réécrire cela sous la forme

$$\int \varphi \, d\mu_{n,\omega} \xrightarrow[n\to\infty]{\text{p.s.}} \int \varphi \, dP_{X_1}.$$

Puisque H est dénombrable, quitte à écarter une réunion dénombrable d'ensembles de probabilité nulle, on obtient

$$\text{p.s. } \forall \varphi \in H, \qquad \int \varphi \, d\mu_{n,\omega} \xrightarrow[n\to\infty]{} \int \varphi \, dP_{X_1}.$$

D'après une proposition du paragraphe précédent, cela suffit pour dire que p.s. $\mu_{n,\omega}$ converge étroitement vers P_{X_1}. □

10.4.2 Le théorème central limite

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi, dans L^1. La loi forte des grands nombres montre que

$$\frac{1}{n} (X_1 + \cdots + X_n) \xrightarrow[n\to\infty]{\text{p.s.}} E[X_1].$$

On cherche alors à savoir à quelle vitesse cette convergence a lieu, c'est-à-dire quel est l'ordre de grandeur de la différence

$$\frac{1}{n} (X_1 + \cdots + X_n) - E[X_1]$$

quand n est grand.

Sous l'hypothèse supplémentaire que les variables X_i sont dans L^2, on devine la réponse en calculant, comme dans la preuve de la loi faible des grands nombres,

$$E[(X_1 + \cdots + X_n - n\,E[X_1])^2] = \mathrm{var}(X_1 + \cdots + X_n) = n\,\mathrm{var}(X_1).$$

Ce calcul indique que la valeur moyenne de $(X_1 + \cdots + X_n - n\,E[X_1])^2$ croît linéairement avec n, donc suggère fortement que l'ordre de grandeur de $X_1 + \cdots + X_n - n\,E[X_1]$ est $\sqrt{n}$, ou encore que l'ordre de grandeur de $\frac{1}{n}(X_1 + \cdots + X_n) - E[X_1]$ est $1/\sqrt{n}$. Le théorème central limite rend ceci plus précis.

Théorème 10.4.2 (Théorème central limite) *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi, dans L^2. Soit $\sigma^2 = \mathrm{var}(X_1)$. Alors,*

$$\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n - n\,E[X_1]) \underset{n\to\infty}{\overset{\text{(loi)}}{\longrightarrow}} \mathcal{N}(0,\sigma^2)$$

où $\mathcal{N}(0,\sigma^2)$ désigne la loi gaussienne centrée de variance σ^2. De manière équivalente, pour tous $a, b \in \bar{\mathbb{R}}$ avec $a < b$,

$$\lim_{n\to\infty} P(X_1 + \cdots + X_n \in [nE[X_1] + a\sqrt{n}, nE[X_1] + b\sqrt{n}]) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp(-\frac{x^2}{2\sigma^2})\,dx.$$

Preuve. La deuxième partie de l'énoncé est une conséquence de la première, compte-tenu de la formulation de la convergence en loi en termes des fonctions de répartition (noter ici que la fonction de répartition de la variable limite est continue). Pour montrer la première partie de l'énoncé, on remarque d'abord qu'on peut supposer $E[X_1] = 0$, quitte à remplacer X_n par $X_n - E[X_n]$. Posons alors

$$Z_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n).$$

La fonction caractéristique de Z_n est

$$\Phi_{Z_n}(\xi) = E\Big[\exp\Big(i\xi(\frac{X_1 + \cdots + X_n}{\sqrt{n}})\Big)\Big] = E\Big[\exp\Big(i\frac{\xi}{\sqrt{n}}X_1\Big)\Big]^n = \Phi_{X_1}(\frac{\xi}{\sqrt{n}})^n,$$

où, dans la seconde égalité, on a utilisé le fait que les v.a. X_i sont indépendantes et de même loi. D'après un résultat du Chapitre 8, on a

$$\Phi_{X_1}(\xi) = 1 + i\xi E[X_1] - \frac{1}{2}\xi^2 E[X_1^2] + o(\xi^2) = 1 - \frac{\sigma^2\xi^2}{2} + o(\xi^2)$$

quand $\xi \to 0$. Pour $\xi \in \mathbb{R}$ fixé, on a donc aussi

$$\Phi_{X_1}(\frac{\xi}{\sqrt{n}}) = 1 - \frac{\sigma^2\xi^2}{2n} + o(\frac{1}{n})$$

quand $n \to \infty$. En combinant avec ce qui précède, on a pour tout $\xi \in \mathbb{R}$,

$$\lim_{n\to\infty} \Phi_{Z_n}(\xi) = \lim_{n\to\infty} (1 - \frac{\sigma^2\xi^2}{2n} + o(\frac{1}{n}))^n = \exp(-\frac{\sigma^2\xi^2}{2}) = \Phi_U(\xi).$$

si U suit la loi $\mathcal{N}(0,\sigma^2)$. Le théorème de Lévy permet maintenant de conclure que Z_n converge en loi vers U, ce qui est le résultat du théorème. $\square$

Cas particulier : Théorème de de Moivre. On suppose que les X_n sont des variables de Bernoulli de paramètre $\frac{1}{2}$ (i.e. $P(X_n = 1) = P(X_n = 0) = \frac{1}{2}$) indépendantes. Alors $S_n = X_1 + \cdots + X_n$ suit une loi binômiale $\mathcal{B}(n, \frac{1}{2})$:

$$P(S_n = k) = C_n^k\, 2^{-n}.$$

Comme $\sigma^2 = 1/4$ dans ce cas particulier, le théorème entraîne que, pour tous $a < b$,

$$2^{-n} \sum_{\frac{n}{2}+a\sqrt{n}\leq k\leq \frac{n}{2}+b\sqrt{n}} C_n^k \underset{n\to\infty}{\longrightarrow} \sqrt{\frac{2}{\pi}} \int_a^b e^{-2x^2}\, dx.$$

Cette dernière convergence peut être vérifiée directement (avec certains efforts) à l'aide de la formule de Stirling. On montre en fait un résultat plus précis de la forme

$$\sqrt{n}\, 2^{-n}\, C_n^k = \sqrt{\frac{2}{\pi}}\, \exp(-\frac{2}{n}(k - \frac{n}{2})^2) + o(1)$$

avec un reste $o(1)$ uniforme quand k varie dans $\{0, 1, \ldots, n\}$.

10.4.3 Extension au cas vectoriel

Supposons maintenant que $(X_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes de même loi à valeurs dans $\mathbb{R}^d$ et intégrables. Alors, on peut appliquer la loi forte des grands nombres coordonnée par coordonnée pour obtenir

$$\frac{1}{n}\,(X_1 + \cdots + X_n) \overset{\text{p.s.}}{\underset{n\to\infty}{\longrightarrow}} E[X_1],$$

où la limite $E[X_1]$ s'interprète évidemment comme le vecteur $(E[X_1^1], \ldots, E[X_1^d])$ si on a écrit $X = (X_1^1, \ldots, X_1^d)$. Supposons de plus que les v.a. X_n sont de carré intégrable. Il n'est pas aussi facile d'obtenir une version multidimensionnelle du théorème central limite : contrairement à ce qui se passe pour la convergence presque sûre, il ne suffit pas pour obtenir la convergence en loi d'une suite de v.a. à valeurs dans $\mathbb{R}^d$ de savoir que chaque suite coordonnée converge en loi (on peut aussi remarquer que la loi de la limite n'est pas déterminée par la connaissance de chacune de ses marginales).

Pour étendre le théorème central limite au cas de v.a. à valeurs dans $\mathbb{R}^d$, nous devons commencer par généraliser la notion de loi gaussienne.

Définition 10.4.1 *Soit C une matrice $d \times d$ à coefficients réels, symétrique positive. Une v.a. X à valeurs dans $\mathbb{R}^d$, de carré intégrable, est appelée vecteur gaussien centré de covariance C si*

$$\forall \xi \in \mathbb{R}^d \ , \quad \Phi_X(\xi) = E[e^{i\xi\cdot X}] = \exp(-\frac{1}{2}\,{}^t\xi C \xi).$$

On dit aussi que X suit la loi $\mathcal{N}(0, C)$.

Remarque. Soit $a \in \mathbb{R}^d$. On dit plus généralement que X suit la loi $\mathcal{N}(a, C)$ si $X - a$ suit la loi $\mathcal{N}(0, C)$.

On a vu dans le Chapitre 8 que si $X = (X^1, \ldots, X^d)$ est une v.a. à valeurs dans $\mathbb{R}^d$ et de carré intégrable, on a le développement limité

$$\Phi_X(\xi) = 1 + i \sum_{j=1}^{d} \xi_j E[X^j] - \frac{1}{2} \sum_{j=1}^{d} \sum_{k=1}^{d} \xi_j \xi_k \, E[X^j X^k] + o(|\xi^2|)$$

quand $\xi \to 0$. On en déduit immédiatement que si X suit la loi $\mathcal{N}(0, C)$ on a $E[X] = 0$ et $K_X = C$.

Proposition 10.4.3 *Soit C une matrice symétrique positive. Il existe un vecteur gaussien centré de covariance C.*

Preuve. Rappelons d'abord (voir la fin du Chapitre 9) qu'une combinaison linéaire de v.a. gaussiennes indépendantes est encore gaussienne.

On pose $A = \sqrt{C}$ de sorte que A est une matrice symétrique positive et $A^2 = C$. Soient ensuite $Y^1, \ldots, Y^d$ d v.a. réelles indépendantes de loi $\mathcal{N}(0, 1)$. Soit Y la v.a. à valeurs dans $\mathbb{R}^d$ dont les coordonnées sont $Y^1, \ldots, Y^d$. Alors, $X = AY$ suit la loi $\mathcal{N}(0, C)$. Pour le voir, considérons $\xi \in \mathbb{R}^d$ et observons que $\xi \cdot X$ est une combinaison linéaire des v.a. $Y^1, \ldots, Y^d$, et est donc une v.a. gaussienne centrée. Précisément, $\xi \cdot X$ suit la loi $\mathcal{N}(0, \sigma^2)$ avec

$$\sigma^2 = E[(\xi \cdot X)^2] = E[{}^t\xi A Y \cdot {}^tY A \xi] = {}^t\xi A \, E[Y\,{}^tY] \, A\xi = {}^t\xi A^2 \xi = {}^t\xi C \xi,$$

en calculant de manière matricielle, et en utilisant le fait que $E[Y\,{}^tY] = Id$ puisque les coordonnées de Y sont des v.a. de loi $\mathcal{N}(0, 1)$ indépendantes. Finalement, grâce à la formule pour la fonction caractéristique d'une v.a. de loi $\mathcal{N}(0, \sigma^2)$, on a pour tout $u > 0$,

$$E[e^{iu\,\xi\cdot X}] = \exp(-\frac{\sigma^2 u^2}{2}) = \exp(-\frac{u^2}{2}\,{}^t\xi C \xi)$$

et en prenant $u = 1$ on a le résultat voulu. □

Remarques. (i) Avec les notations de la preuve ci-dessus, Y suit la loi $\mathcal{N}(0, Id)$.
(ii) Une v.a. X à valeurs dans $\mathbb{R}^d$ est un vecteur gaussien centré si et seulement si toute combinaison linéaire de ses composantes est gaussienne centrée : en effet on a alors $E[e^{i\xi\cdot X}] = \exp(-\frac{1}{2}\,E[(\xi \cdot X)^2]) = \exp(-\frac{1}{2}\,{}^t\xi K_X \xi)$.

Exercice. Soit X un vecteur gaussien centré. Montrer que X a une densité si et seulement si K_X est non dégénérée, et calculer alors la densité de X.

Théorème 10.4.4 (Théorème central limite vectoriel) *Soit $(X_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes de même loi à valeurs dans $\mathbb{R}^d$, de carré intégrable. Alors,*

$$\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n - n\,E[X_1]) \xrightarrow[n\to\infty]{\text{(loi)}} \mathcal{N}(0, K_{X_1})$$

Preuve. C'est la même que dans le cas réel. On peut supposer $E[X_1] = 0$. Ensuite, pour tout $\xi \in \mathbb{R}^d$,

$$E\Big[\exp\Big(i\xi\cdot(\frac{X_1+\cdots+X_n}{\sqrt{n}})\Big)\Big] = E\Big[\exp\Big(i\frac{\xi}{\sqrt{n}}\cdot X_1\Big)\Big]^n = \Phi_{X_1}(\frac{\xi}{\sqrt{n}})^n.$$

D'autre part, on sait que

$$\Phi_{X_1}(\frac{\xi}{\sqrt{n}}) = 1 - \frac{1}{2n}\,{}^t\xi K_{X_1}\xi + o(\frac{1}{n}).$$

On conclut que

$$\lim_{n\to\infty} E\Big[\exp\Big(i\xi\cdot(\frac{X_1+\cdots+X_n}{\sqrt{n}})\Big)\Big] = \exp(-\frac{1}{2}\,{}^t\xi K_{X_1}\xi),$$

d'où le résultat grâce au théorème de Lévy. □

Chapitre 11

Conditionnement

Ce chapitre est consacré à la construction et aux propriétés de l'espérance conditionnelle. Intuitivement, l'espérance conditionnelle d'une variable aléatoire réelle donnée par rapport à une sous-tribu est la variable aléatoire mesurable pour cette sous-tribu qui est la "plus proche" de la variable aléatoire donnée. Pour de nombreux problèmes concrets (prédiction, observation incomplète, etc.) il est important de pouvoir estimer une variable aléatoire sur laquelle on n'a qu'une information partielle, et l'on comprend dès lors l'importance de la notion d'espérance conditionnelle. La définition axiomatique de cette notion (dans laquelle la "propriété caractéristique" joue un rôle essentiel) est motivée par le cas discret traité dans le premier paragraphe. Le calcul explicite des espérances conditionnelles, qui est en général un problème difficile, est illustré sur plusieurs cas, dont le cas gaussien particulièrement important pour les applications. La notion de loi conditionnelle, utile dans ce cours d'un point de vue conceptuel surtout, est introduite à la fin du chapitre.

11.1 Conditionnement discret

Comme dans les chapitres précédents on se place sur un espace de probabilité $(\Omega, \mathcal{A}, P)$. Soit $B \in \mathcal{A}$ un événement tel que $P(B) > 0$. On peut définir une nouvelle probabilité sur $(\Omega, \mathcal{A})$, appelée probabilité conditionnelle sachant B, en posant pour tout $A \in \mathcal{A}$,

$$P(A\,|\,B) = \frac{P(A \cap B)}{P(B)}.$$

De même, pour toute v.a. $X \geq 0$, ou pour $X \in L^1(\Omega, \mathcal{A}, P)$, l'espérance conditionnelle de X sachant B est définie par

$$E[X\,|\,B] = \frac{E[X\,1_B]}{P(B)}.$$

Cette quantité est aussi l'espérance de X sous la probabilité $P(\cdot \mid B)$, et elle s'interprète comme la valeur moyenne de X quand B est réalisé.

Nous cherchons ensuite à définir l'espérance conditionnelle sachant une variable aléatoire (et non plus sachant un événement). Considérons une v.a. Y à valeurs dans un espace E

dénombrable. Soit $E' = \{y \in E : P(Y = y) > 0\}$. Pour tout $y \in E'$, et pour toute v.a. $X \in L^1(\Omega, \mathcal{A}, P)$, on peut définir, comme cas particulier de ce qui précède,

$$E[X \,|\, Y = y] = \frac{E[X\, 1_{\{Y=y\}}]}{P(Y = y)}.$$

Définition 11.1.1 *Soit $X \in L^1(\Omega, \mathcal{A}, P)$. L'espérance conditionnelle de X sachant Y est la variable aléatoire réelle définie par*

$$E[X \,|\, Y] = \varphi(Y),$$

où la fonction $\varphi : E \longrightarrow \mathbb{R}$ est donnée par

$$\varphi(y) = \begin{cases} E[X \,|\, Y = y] & \text{si } y \in E', \\ 0 & \text{si } y \in E \backslash E'. \end{cases}$$

Remarque. Le choix de la valeur de φ sur $E \backslash E'$ est arbitraire : de toute façon ce choix n'influence la définition de $E[X \,|\, Y]$ que sur un ensemble de probabilité nulle, puisque

$$P(Y \in E \backslash E') = \sum_{y \in E \backslash E'} P(Y = y) = 0.$$

On pourrait changer la définition de φ sur $E \backslash E'$ et cela donnerait la même v.a. $E[X \,|\, Y]$ à un ensemble de mesure nulle près. Dans les situations plus générales que nous rencontrerons plus tard, les espérances conditionnelles (sachant une v.a. ou une tribu) seront toujours définies à un ensemble de probabilité nulle près.

En comparant avec le conditionnement par rapport à un événement, on observe que l'espérance conditionnelle $E[X \,|\, Y]$ est maintenant une variable aléatoire : c'est la v.a. qui donne la valeur moyenne de X quand on connait Y : p.s.,

$$E[X \,|\, Y](\omega) = E[X \,|\, Y = y] \,, \qquad \text{si } Y(\omega) = y.$$

Remarquons aussi que $E[X \,|\, Y]$ est une fonction de Y donc une v.a. $\sigma(Y)$-mesurable. Dans un sens qui sera précisé plus loin, c'est la meilleure approximation de X par une fonction de Y.

Exemple. *Lancer d'un dé.* On prend $\Omega = \{1, 2, \ldots, 6\}$ et $P(\{\omega\}) = \frac{1}{6}$ pour tout $\omega \in \Omega$. Soient

$$Y(\omega) = \begin{cases} 1 & \text{si } \omega \text{ est impair,} \\ 0 & \text{si } \omega \text{ est pair,} \end{cases}$$

et $X(\omega) = \omega$. Alors,

$$E[X \,|\, Y](\omega) = \begin{cases} 3 & \text{si } \omega \in \{1, 3, 5\}, \\ 4 & \text{si } \omega \in \{2, 4, 6\}. \end{cases}$$

Proposition 11.1.1 *On a $E[|E[X \,|\, Y]|] \le E[|X|]$. En particulier, $E[X \,|\, Y] \in L^1(\Omega, \mathcal{A}, P)$. De plus, pour toute v.a. Z $\sigma(Y)$-mesurable bornée,*

$$E[ZX] = E[Z\, E[X \,|\, Y]].$$

Preuve. D'après la définition de l'espérance conditionnelle $E[X\,|\,Y]$, on a

$$E[|E[X\,|\,Y]|] = \sum_{y\in E'} P(Y=y)\,\frac{|E[X\,1_{\{Y=y\}}]|}{P(Y=y)} \leq \sum_{y\in E} E[|X|\,1_{\{Y=y\}}] = E[|X|].$$

Pour la dernière assertion, on utilise le fait qu'on peut écrire $Z = \psi(Y)$, avec une fonction ψ bornée. Alors,

$$E[\psi(Y)\,E[X\,|\,Y]] = \sum_{y\in E} \psi(y)\,E[X\,1_{\{Y=y\}}] = \sum_{y\in E} E[\psi(Y)X\,1_{\{Y=y\}}] = E[\psi(Y)X].$$

□

Conséquence. Si Y' est une autre v.a. discrète telle que $\sigma(Y) = \sigma(Y')$, on a

$$E[X\,|\,Y] = E[X\,|\,Y'] \qquad \text{p.s.}$$

En effet, en appliquant la proposition avec $Z = 1_{\{E[X|Y]>E[X|Y']\}}$, qui est bien mesurable pour $\sigma(Y) = \sigma(Y')$ puisque $E[X\,|\,Y]$ et $E[X\,|\,Y']$ le sont, on trouve

$$E[1_{\{E[X|Y]>E[X|Y']\}}(E[X\,|\,Y] - E[X\,|\,Y'])] = 0$$

d'où $E[X\,|\,Y] \leq E[X\,|\,Y']$ p.s., et on obtient de même l'autre inégalité. Cela montre aussi que la dernière propriété de la proposition caractérise $E[X\,|\,Y]$ parmi les v.a. $\sigma(Y)$-mesurables et intégrables.

L'observation précédente conduit à dire que la "bonne" notion de conditionnement est la notion de conditionnement par rapport à une tribu. C'est cette notion que nous allons développer dans les paragraphes suivants en nous basant sur la propriété de la proposition ci-dessus.

11.2 La définition de l'espérance conditionnelle

11.2.1 Cas des variables intégrables

Théorème et définition 11.2.1 *Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$, et soit $X \in L^1(\Omega, \mathcal{A}, P)$. Il existe alors une unique variable aléatoire dans $L^1(\Omega, \mathcal{B}, P)$, notée $E[X\,|\,\mathcal{B}]$, telle que*

$$\forall B \in \mathcal{B}\ , \qquad E[X\,1_B] = E[E[X\,|\,\mathcal{B}]\,1_B]. \tag{11.1}$$

On a plus généralement, pour toute variable aléatoire Z $\mathcal{B}$-mesurable bornée

$$E[X\,Z] = E[E[X\,|\,\mathcal{B}]\,Z]. \tag{11.2}$$

Si $X \geq 0$ on a aussi $E[X\,|\,\mathcal{B}] \geq 0$.

Le point crucial est le fait que $E[X\,|\,\mathcal{B}]$ est mesurable pour la tribu $\mathcal{B}$. L'une ou l'autre des propriétés (11.1) et (11.2) caractérise l'espérance conditionnelle $E[X\,|\,\mathcal{B}]$ dans la classe

des v.a. de $L^1(\Omega, \mathcal{B}, P)$. Dans la suite nous ferons référence à l'une ou l'autre comme à la *propriété caractéristique* de l'espérance conditionnelle.

Dans le cas particulier où la tribu $\mathcal{B}$ est engendrée par une variable aléatoire Y, on écrira indifféremment

$$E[X \,|\, \mathcal{B}] = E[X \,|\, \sigma(Y)] = E[X \,|\, Y].$$

Cette notation est cohérente avec le cas discret traité dans la partie précédente : comparer (11.2) et la proposition ci-dessus.

Preuve. Commençons par l'unicité. Soient X' et X'' deux v.a. dans $L^1(\Omega, \mathcal{B}, P)$ telles que

$$\forall B \in \mathcal{B}\ , \qquad E[X' \, 1_B] = E[X \, 1_B] = E[X'' \, 1_B].$$

En prenant $B = \{X' > X''\}$ (qui est bien $\mathcal{B}$-mesurable puisque X' et X'' le sont), on trouve

$$E[(X' - X'')1_{\{X' > X''\}}] = 0$$

d'où $X' \leq X''$ p.s., et de même $X' \geq X''$ p.s.

Pour l'existence, supposons d'abord $X \geq 0$, et soit Q la mesure finie sur $(\Omega, \mathcal{B})$ définie par

$$\forall B \in \mathcal{B}\ , \qquad Q(B) = E[X \, 1_B].$$

Alors, si on voit aussi P comme une mesure de probabilité sur $(\Omega, \mathcal{B})$, il est immédiat qu'on a $Q \ll P$. Le théorème de Radon-Nikodym, appliqué sur l'espace mesurable $(\Omega, \mathcal{B})$, assure donc l'existence d'une v.a. $\widetilde{X}$ $\mathcal{B}$-mesurable positive telle que

$$\forall B \in \mathcal{B}\ , \qquad E[X \, 1_B] = Q(B) = E[\widetilde{X} \, 1_B].$$

En prenant $B = \Omega$, on voit que $E[\widetilde{X}] = E[X] < \infty$, donc $X \in L^1(\Omega, \mathcal{B}, P)$. Finalement, $E[X \,|\, \mathcal{B}] = \widetilde{X}$ vérifie la propriété de l'énoncé. Lorsque X est de signe quelconque, il suffit de prendre

$$E[X \,|\, \mathcal{B}] = E[X^+ \,|\, \mathcal{B}] - E[X^- \,|\, \mathcal{B}].$$

Enfin, le passage de (11.1) à (11.2) se fait en utilisant l'approximation usuelle des fonctions mesurables par des fonctions étagées. $\square$

Exemple. Prenons $\Omega =]0,1]$, $\mathcal{A} = \mathcal{B}(]0,1])$ et $P(d\omega) = d\omega$. Soit $\mathcal{B}$ la tribu engendrée par les intervalles $]\frac{i-1}{n}, \frac{i}{n}]$, $i \in \{1, \ldots, n\}$, où $n \geq 1$ est fixé. Un élément f de $L^1(\Omega, \mathcal{A}, P)$ est une fonction mesurable $f :]0,1] \longrightarrow \mathbb{R}$ telle que $\int_0^1 |f(\omega)| d\omega < \infty$. Alors on vérifie très facilement que

$$E[f \,|\, \mathcal{B}] = \sum_{i=1}^{n} f_i \, 1_{]\frac{i-1}{n}, \frac{i}{n}]}\ ,$$

où $f_i = n \int_{(i-1)/n}^{i/n} f(\omega) d\omega$ est la moyenne de f sur $]\frac{i-1}{n}, \frac{i}{n}]$.

Propriétés de l'espérance conditionnelle.

(a) Si X est $\mathcal{B}$-mesurable, $E[X \,|\, \mathcal{B}] = X$.

(b) L'application $X \longrightarrow E[X \,|\, \mathcal{B}]$ est linéaire.

(c) $E[E[X\,|\,\mathcal{B}]] = E[X]$.

(d) $|E[X\,|\,\mathcal{B}]| \leq E[|X|\,|\,\mathcal{B}]$ p.s., et en conséquence $E[|E[X\,|\,\mathcal{B}]|] \leq E[|X|]$.

(e) $X \geq X' \Rightarrow E[X\,|\,\mathcal{B}] \geq E[X'\,|\,\mathcal{B}]$ p.s.

Preuve. (a) découle immédiatement de l'unicité dans le théorème ci-dessus. Il en va de même pour (b) en observant que, si $X, X' \in L^1(\Omega, \mathcal{A}, P)$ et $\alpha, \alpha' \in \mathbb{R}$, la v.a.

$$\alpha\, E[X\,|\,\mathcal{B}] + \alpha' E[X'\,|\,\mathcal{B}]$$

satisfait la propriété caractéristique (11.1) pour $\alpha X + \alpha' X'$. La propriété (c) est le cas particulier $B = \Omega$ dans (11.1). Pour (d), rappelons que si $X \geq 0$ on a $E[X\,|\,\mathcal{B}] \geq 0$. Cela entraîne

$$|E[X\,|\,\mathcal{B}]| = |E[X^+\,|\,\mathcal{B}] - E[X^-\,|\,\mathcal{B}]| \leq E[X^+\,|\,\mathcal{B}]] + E[X^-\,|\,\mathcal{B}] = E[|X|\,|\,\mathcal{B}].$$

Enfin, (e) est immédiat par linéarité. □

11.2.2 Cas des variables positives

Théorème 11.2.2 *Soit X une variable aléatoire à valeurs dans $[0, \infty]$. La formule*

$$E[X\,|\,\mathcal{B}] = \lim_{n\to\infty} \uparrow E[X \wedge n\,|\,\mathcal{B}] \qquad p.s.$$

définit une variable aléatoire à valeurs dans $[0, \infty]$, qui est caractérisée (à un ensemble de probabilité nulle près) par la propriété suivante : pour toute variable aléatoire Z $\mathcal{B}$-mesurable positive,

$$E[XZ] = E[E[X\,|\,\mathcal{B}]Z]. \tag{11.3}$$

Dans le cas où X est aussi intégrable, en comparant la dernière propriété du théorème avec (11.1), on voit immédiatement que l'on retrouve la même définition de $E[X \mid \mathcal{B}]$ que dans le paragraphe ci-dessus. De même que dans le cas des variables intégrables, la propriété (11.3) sera appelée propriété caractéristique de l'espérance conditionnelle.

Preuve. La croissance de la limite dans la définition de $E[X \mid \mathcal{B}]$ découle de la propriété (e) ci-dessus. Ensuite, si Z est $\mathcal{B}$-mesurable positive, le théorème de convergence monotone entraîne que

$$E[E[X\,|\,\mathcal{B}]Z] = \lim_{n\to\infty} E[E[X \wedge n\,|\,\mathcal{B}](Z \wedge n)] = \lim_{n\to\infty} E[(X \wedge n)(Z \wedge n)] = E[XZ].$$

Il reste à établir l'unicité. Soient donc X' et X'' deux variables aléatoires $\mathcal{B}$-mesurables à valeurs dans $[0, \infty]$ telles que

$$E[X'Z] = E[X''Z]$$

pour toute v.a. Z $\mathcal{B}$-mesurable positive. Prenons

$$Z = 1_{\{X' \leq a < b \leq X''\}}$$

où on a fixé $a, b \in \mathbb{Q}_+$, avec $a < b$. Il vient

$$a\, P(X' \le a < b \le X'') \ge b\, P(X' \le a < b \le X'')$$

ce qui n'est possible que si $P(X' \le a < b \le X'') = 0$. On a donc

$$P\Big(\bigcup_{\substack{a,b \in \mathbb{Q}_+ \\ a<b}} \{X' \le a < b \le X''\} \Big) = 0$$

ce qui entraîne $X' \ge X''$ p.s. Par un raisonnement symétrique on a aussi $X'' \ge X'$ p.s. $\square$

Remarque. On peut avoir $X < \infty$ p.s. et simultanément $P(E[X \mid \mathcal{B}] = \infty) > 0$. Par exemple, si $\mathcal{B} = \{\varnothing, \Omega\}$, on vérifie aisément que $E[X \mid \mathcal{B}] = E[X]$, qui peut bien sûr être infini pour des v.a. X finies p.s. Pour donner un exemple moins trivial, reprenons le cas où $\Omega =]0,1]$, $\mathcal{B} = \sigma(]\frac{i-1}{n}, \frac{i}{n}]; i \in \{1, \dots, n\})$ et $P(d\omega) = d\omega$. Alors, si $X(\omega) = \frac{1}{\omega}$, on a

$$E[X \,|\, \mathcal{B}] = \infty\, 1_{]0,\frac{1}{n}]} + \sum_{i=2}^{n} n \,\log(\frac{i}{i-1})\, 1_{]\frac{i-1}{n},\frac{i}{n}]}.$$

Propriétés.

(a) Si X et X' sont des v.a. positives et $a, b \ge 0$,

$$E[aX + bX' \,|\, \mathcal{B}] = a\, E[X \,|\, \mathcal{B}] + b\, E[X' \,|\, \mathcal{B}].$$

(b) Si X est $\mathcal{B}$-mesurable, $E[X \,|\, \mathcal{B}] = X$.

(c) Si (X_n) est une suite croissante de v.a. positives, et $X = \lim \uparrow X_n$,

$$E[X \,|\, \mathcal{B}] = \lim_{n\to\infty} \uparrow E[X_n \,|\, \mathcal{B}] \;, \qquad \text{p.s.}$$

(d) Si (X_n) est une suite de v.a. positives,

$$E[\liminf X_n \,|\, \mathcal{B}] \le \liminf E[X_n \,|\, \mathcal{B}] \;, \qquad \text{p.s.}$$

(e) Soit (X_n) une suite de v.a. intégrables convergeant p.s. vers X. Supposons qu'il existe une v.a. positive Z telle que $|X_n| \le Z$ p.s. pour tout n, et $E[Z] < \infty$. Alors,

$$E[X \,|\, \mathcal{B}] = \lim_{n\to\infty} E[X_n \,|\, \mathcal{B}] \;, \qquad \text{p.s. et dans } L^1.$$

(f) Si f est convexe positive, et si $X \in L^1$,

$$E[f(X) \,|\, \mathcal{B}] \ge f(E[X \,|\, \mathcal{B}]).$$

Remarque. La mention "p.s." devrait figurer dans chaque énoncé impliquant une espérance conditionnelle, puisque celle-ci n'est définie qu'à un ensemble de probabilité nulle près. Le plus souvent cependant, cette mention est sous-entendue, comme dans (a),(b) et (f) ci-dessus.

Preuve. (a) et (b) sont faciles en utilisant la caractérisation de $E[X \mid \mathcal{B}]$ donnée dans le théorème.

(c) Il découle de (a) que si $X_1 \geq X_2 \geq 0$ on a $E[X_1 | \mathcal{B}] \geq E[X_2 | \mathcal{B}]$. Sous les hypothèses de (c), on peut donc poser $X' = \lim \uparrow E[X_n | \mathcal{B}]$, qui est une v.a. $\mathcal{B}$-mesurable à valeurs dans $[0, \infty]$. On a alors, pour toute v.a. Z $\mathcal{B}$-mesurable positive,

$$E[ZX'] = \lim \uparrow E[Z\, E[X_n | \mathcal{B}]] = \lim \uparrow E[Z\, X_n] = E[ZX]$$

ce qui d'après la caractérisation du théorème entraîne $X' = E[X | \mathcal{B}]$.

(d) On écrit, en utilisant (c),

$$\begin{aligned} E[\liminf X_n | \mathcal{B}] &= E\Big[\lim_{k\uparrow\infty} \uparrow \Big(\inf_{n\geq k} X_n\Big) \Big| \mathcal{B}\Big] \\ &= \lim_{k\uparrow\infty} \uparrow E\Big[\inf_{n\geq k} X_n \Big| \mathcal{B}\Big] \\ &\leq \lim_{k\uparrow\infty} \Big(\inf_{n\geq k} E[X_n \mid \mathcal{B}]\Big) \\ &= \liminf E[X_n | \mathcal{B}]. \end{aligned}$$

(e) Il suffit d'appliquer (d) deux fois :

$$\begin{aligned} E[\liminf (Z - X_n) | \mathcal{B}] &\leq E[Z | \mathcal{B}] - \limsup E[X_n | \mathcal{B}] \\ E[\liminf (Z + X_n) | \mathcal{B}] &\leq E[Z | \mathcal{B}] + \liminf E[X_n | \mathcal{B}] \end{aligned}$$

ce qui conduit à

$$E[X | \mathcal{B}] \leq \liminf E[X_n | \mathcal{B}] \leq \limsup E[X_n | \mathcal{B}] \leq E[X | \mathcal{B}],$$

d'où la convergence p.s. recherchée. La convergence L^1 est maintenant une conséquence du théorème de convergence dominée, puisque $|E[X_n \mid \mathcal{B}]| \leq E[|X_n| \mid \mathcal{B}] \leq E[Z \mid \mathcal{B}]$ et $E[E[Z | \mathcal{B}]] = E[Z] < \infty$.

(f) Notons

$$E_f = \{(a, b) \in \mathbb{R}^2 : \forall x \in \mathbb{R},\ f(x) \geq ax + b\}.$$

Alors, il est facile de vérifier que

$$\forall x \in \mathbb{R}^2\ , \qquad f(x) = \sup_{(a,b)\in E_f} (ax + b) = \sup_{(a,b)\in E_f \cap \mathbb{Q}^2} (ax + b).$$

En utilisant le fait que $\mathbb{Q}^2$ est dénombrable, on en déduit que p.s.

$$E[f(X) | \mathcal{B}] = E\Big[\sup_{(a,b)\in E_f\cap\mathbb{Q}^2} (aX + b) \Big| \mathcal{B}\Big] \geq \sup_{(a,b)\in E_f\cap\mathbb{Q}^2} E[aX + b | \mathcal{B}] = f(E[X | \mathcal{B}]).$$

□

Remarque. Par analogie avec la formule $P(A) = E[1_A]$, on écrira souvent pour $A \in \mathcal{A}$,

$$P(A | \mathcal{B}) := E[1_A | \mathcal{B}].$$

Prendre garde cependant que $P(A | \mathcal{B})$ ainsi définie est une *variable aléatoire*.

11.2.3 Le cas particulier des variables de carré intégrable

Dans le cas où X est de carré intégrable, il existe une autre interprétation remarquable de $E[X\,|\,\mathcal{B}]$. Avant d'énoncer le résultat, observons que $L^2(\Omega, \mathcal{B}, P)$ s'identifie à un sous-espace fermé de $L^2(\Omega, \mathcal{A}, P)$, à savoir l'espace des éléments de $L^2(\Omega, \mathcal{A}, P)$ dont un représentant au moins est $\mathcal{B}$-mesurable.

Théorème 11.2.3 *Si $X \in L^2(\Omega, \mathcal{A}, P)$, alors $E[X \mid \mathcal{B}]$ est la projection orthogonale de X sur $L^2(\Omega, \mathcal{B}, P)$.*

Preuve. La propriété (f) ci-dessus montre que $E[X\,|\,\mathcal{B}]^2 \leq E[X^2\,|\,\mathcal{B}]$ p.s. Cela entraîne que $E[E[X\,|\,\mathcal{B}]^2] \leq E[X^2] < \infty$, et donc la v.a. $E[X\,|\,\mathcal{B}]$ est dans $L^2(\Omega, \mathcal{B}, P)$.

Par ailleurs, pour toute v.a. Z $\mathcal{B}$-mesurable bornée,

$$E[Z(X - E[X\,|\,\mathcal{B}])] = E[ZX] - E[ZE[X\,|\,\mathcal{B}]] = 0,$$

toujours d'après la propriété caractéristique de $E[X\,|\,\mathcal{B}]$. Donc $X - E[X\,|\,\mathcal{B}]$ est orthogonal à toutes les v.a. bornées $\mathcal{B}$-mesurables, et par un argument de densité, $X - E[X \mid \mathcal{B}]$ est orthogonal à $L^2(\Omega, \mathcal{B}, P)$. Le résultat annoncé en découle. □

On peut utiliser le théorème précédent pour donner une autre construction de l'espérance conditionnelle, évitant le recours au théorème de Radon-Nikodym, en commençant par le cas des v.a. de carré intégrable. Observons aussi que ce théorème donne une interprétation intéressante de l'espérance conditionnelle : si X est de carré intégrable, $E[X \mid \mathcal{B}]$ est la meilleure (au sens de la norme L^2) approximation de X par une v.a. $\mathcal{B}$-mesurable.

11.3 Propriétés spécifiques de l'espérance conditionnelle

Les propriétés établies ci-dessus sont analogues aux propriétés de l'espérance (ou de l'intégrale de fonctions mesurables). Nous établissons dans ce paragraphe des propriétés plus particulières à l'espérance conditionnelle.

Proposition 11.3.1 *Soit X une variable aléatoire réelle, et soit Y une variable aléatoire $\mathcal{B}$-mesurable. Alors,*

$$E[YX\,|\,\mathcal{B}] = Y\,E[X\,|\,\mathcal{B}]$$

dès que les espérances conditionnelles sont bien définies, c'est-à-dire si X et Y sont positives, ou si X et $YX \in L^1$.

Preuve. Supposons $X \geq 0$ et $Y \geq 0$. Alors, pour toute v.a. Z $\mathcal{B}$-mesurable positive,

$$E[Z(YE[X\,|\,\mathcal{B}])] = E[(ZY)E[X\,|\,\mathcal{B}]] = E[ZYX].$$

Puisque $YE[X\,|\,\mathcal{B}]$ est une v.a. $\mathcal{B}$-mesurable positive, cette égalité suffit pour conclure que $YE[X\,|\,\mathcal{B}] = E[YX\,|\,\mathcal{B}]$.

Dans le cas où X et YX sont intégrables, on obtient le résultat en décomposant $X = X^+ - X^-$ et $Y = Y^+ - Y^-$. □

Proposition 11.3.2 *Soient $\mathcal{B}_1$ et $\mathcal{B}_2$ deux sous-tribus de $\mathcal{A}$ telles que $\mathcal{B}_1 \subset \mathcal{B}_2$. Alors, pour toute variable aléatoire X positive ou intégrable,*

$$E[E[X\,|\,\mathcal{B}_2]\,|\,\mathcal{B}_1] = E[X\,|\,\mathcal{B}_1].$$

Remarque. On a aussi $E[E[X\,|\,\mathcal{B}_1]\,|\,\mathcal{B}_2] = E[X\,|\,\mathcal{B}_1]$ sous les mêmes hypothèses, mais cela est évident puisque $E[X\,|\,\mathcal{B}_1]$ est $\mathcal{B}_2$-mesurable.
Preuve. Traitons le cas où $X \geq 0$. Soit Z une v.a. $\mathcal{B}_1$-mesurable positive. Alors, puisque Z est aussi $\mathcal{B}_2$-mesurable,

$$E[Z\,E[E[X\,|\,\mathcal{B}_2]\,|\,\mathcal{B}_1]] = E[Z\,E[X\,|\,\mathcal{B}_2]] = E[ZX].$$

Cela suffit pour établir l'égalité annoncée. □

Théorème 11.3.3 *Deux sous-tribus $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes si et seulement si, pour toute v.a. X $\mathcal{B}_2$-mesurable positive (ou pour toute v.a. $X \in L^1(\Omega, \mathcal{B}_2, P)$, ou bien pour toute v.a. X de la forme $X = 1_A$, avec $A \in \mathcal{B}_2$), on a*

$$E[X\,|\,\mathcal{B}_1] = E[X].$$

Preuve. Supposons d'abord que $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes. Alors, si X est une v.a. $\mathcal{B}_2$-mesurable positive, on a pour toute v.a. Z $\mathcal{B}_1$-mesurable positive,

$$E[ZX] = E[Z]E[X] = E[Z\,E[X]],$$

et donc la v.a. constante $E[X]$ satisfait la propriété caractéristique de l'espérance conditionnelle $E[X\,|\,\mathcal{B}_1]$. Dans le cas où X est intégrable, il suffit d'utiliser la linéarité de l'espérance conditionnelle.

Supposons inversement que

$$\forall A \in \mathcal{B}_2\ , \qquad E[1_A\,|\,\mathcal{B}_1] = E[1_A] = P(A).$$

Alors, pour tout $B \in \mathcal{B}_1$,

$$P(A \cap B) = E[1_A 1_B] = E[E[1_A\,|\,\mathcal{B}_1]\,1_B] = E[P(A)\,1_B] = P(A)P(B)$$

ce qui montre que les tribus $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes. □

Remarque. Soient X et Y deux v.a. réelles. Puisque les v.a. mesurables par rapport à la tribu $\sigma(X)$ sont les fonctions de X, le théorème précédent montre que X et Y sont indépendantes si et seulement si

$$E[h(X)\,|\,Y] = E[h(X)]$$

pour toute fonction borélienne h telle que $E[|h(X)|] < \infty$ (rappelons que $E[h(X)\mid Y] = E[h(X)\,|\,\sigma(Y)]$). Si X est intégrable on a donc en particulier

$$E[X\,|\,Y] = E[X].$$

Cependant cette dernière propriété seule ne suffit pas pour donner l'indépendance de X et Y. Pour s'en convaincre, il suffit de traiter le cas où X suit une loi $\mathcal{N}(0,1)$, et $Y = |X|$. Alors, toute v.a. Z $\sigma(Y)$-mesurable bornée s'écrit $Z = g(Y)$, avec une fonction g bornée, et donc

$$E[ZX] = E[g(|X|)X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dy\, e^{-y^2/2}\, g(|y|)y = 0,$$

ce qui montre que $E[X\,|\,Y] = 0 = E[X]$, alors que X et Y ne sont bien sûr pas indépendantes.

Nous énonçons maintenant un autre théorème reliant espérance conditionnelle et indépendance, qui est très souvent utile pour les calculs explicites d'espérance conditionnelle.

Théorème 11.3.4 *Soient X et Y deux variables aléatoires à valeurs respectivement dans les espaces mesurables E et F. Supposons que X est indépendante de $\mathcal{B}$ et que Y est $\mathcal{B}$-mesurable. Alors, pour toute fonction mesurable $g : E \times F \longrightarrow \mathbb{R}_+$,*

$$E[g(X,Y) \mid \mathcal{B}] = \int g(x,Y)\, P_X(dx),$$

où P_X désigne la loi de X. Le terme de droite est la composée de la variable aléatoire Y par l'application $\Phi : y \longrightarrow \int g(x,y)\, P_X(dx)$ (Φ est mesurable grâce au théorème de Fubini).

Remarque. De manière informelle on peut expliquer le théorème de la manière suivante. Si on conditionne par rapport à la sous-tribu $\mathcal{B}$, la v.a. Y, qui est $\mathcal{B}$-mesurable, se comporte comme une constante et comme par ailleurs la connaissance de $\mathcal{B}$ ne donne aucune information sur X la meilleure approximation de $g(X,Y)$ est obtenue en intégrant $g(\cdot,Y)$ par rapport à la loi de X.

Preuve. Il suffit de montrer que pour toute v.a. Z $\mathcal{B}$-mesurable positive,

$$E[g(X,Y)Z] = E[\Phi(Y)Z].$$

Notons $P_{(X,Y,Z)}$ la loi du triplet (X,Y,Z), qui est une mesure de probabilité sur $E \times F \times \mathbb{R}_+$. Comme X est indépendante de (Y,Z), on a

$$P_{(X,Y,Z)} = P_X \otimes P_{(Y,Z)}$$

et donc, en utilisant le théorème de Fubini,

$$\begin{aligned}
E[g(X,Y)Z] &= \int g(x,y)z\, P_{(X,Y,Z)}(dxdydz) \\
&= \int g(x,y)z\, P_X(dx)P_{(Y,Z)}(dydz) \\
&= \int_{F\times\mathbb{R}_+} z\Big(\int_E g(x,y)P_X(dx)\Big)P_{(Y,Z)}(dydz) \\
&= \int_{F\times\mathbb{R}_+} z\Phi(y)\, P_{(Y,Z)}(dydz) \\
&= E[\Phi(Y)Z]
\end{aligned}$$

ce qui était le résultat recherché. □

11.4 Calculs d'espérance conditionnelle

11.4.1 Conditionnement discret

Soit Y une v.a. à valeurs dans un espace dénombrable E, et soit $X \in L^1(\Omega, \mathcal{A}, P)$. Alors on a déjà vu que

$$E[X \,|\, Y] = \varphi(Y)$$

où

$$\varphi(y) = \frac{E[X \, 1_{\{Y=y\}}]}{P(Y=y)}$$

pour tout $y \in E$ tel que $P(Y = y) > 0$ (et $\varphi(y)$ peut être choisie de manière arbitraire lorsque $P(Y = y) = 0$).

11.4.2 Cas des variables à densité

Soient X et Y deux v.a. à valeurs respectivement dans $\mathbb{R}^m$ et dans $\mathbb{R}^n$. Supposons que le couple (X, Y) a pour densité $p(x, y)$: pour toute fonction borélienne $f : \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R}_+$,

$$E[f(X,Y)] = \int_{\mathbb{R}^m \times \mathbb{R}^n} f(x,y)\, p(x,y)\, dxdy.$$

Alors la densité de Y est la fonction

$$q(y) = \int_{\mathbb{R}^m} p(x,y)\, dx$$

(en toute rigueur il faut prendre $q(y) = 0$ pour les valeurs de y telles que $\int p(x, y)\, dx = \infty$, qui forment un ensemble de mesure nulle; nous négligerons cependant ce point de détail dans les calculs qui suivent).

Soit maintenant $h : \mathbb{R}^m \longrightarrow \mathbb{R}_+$ une fonction mesurable. Alors on calcule $E[h(X) \,|\, Y]$ de la façon suivante. Pour toute fonction $g : \mathbb{R}^n \longrightarrow \mathbb{R}_+$ borélienne, on a

$$\begin{aligned}
E[h(X)g(Y)] &= \int_{\mathbb{R}^m \times \mathbb{R}^n} h(x)\, g(y)\, p(x,y)\, dxdy \\
&= \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^m} h(x)\, p(x,y)\, dx \Big) g(y)\, dy \\
&= \int_{\mathbb{R}^n} \Big(\frac{\int_{\mathbb{R}^m} h(x)\, p(x,y)\, dx}{q(y)} \Big) g(y)\, q(y) 1_{\{q(y)>0\}}\, dy \\
&= \int_{\mathbb{R}^n} \varphi(y)\, g(y)\, q(y) 1_{\{q(y)>0\}}\, dy \\
&= E[\varphi(Y)\, g(Y)],
\end{aligned}$$

où on a posé

$$\varphi(y) = \begin{cases} \dfrac{1}{q(y)} \displaystyle\int_{\mathbb{R}^m} h(x)\, p(x,y)\, dx & \text{si } q(y) > 0, \\ h(0) & \text{si } q(y) = 0 \end{cases}$$

(la valeur de $\varphi(y)$ lorsque $q(y) = 0$ est arbitraire : le choix de la valeur $h(0)$ sera commode dans l'énoncé qui suit). Dans le calcul qui précède, on a utilisé implicitement le fait que si $q(y) = 0$ on a $p(x, y) = 0$ dx p.p., et donc $\int h(x)\, p(x, y)\, dx = 0$.

Il découle du calcul ci-dessus et de la caractérisation de l'espérance conditionnelle que

$$E[h(X)\,|\,Y] = \varphi(Y).$$

Nous réénonçons ce résultat sous une forme un peu différente.

Proposition 11.4.1 *Pour tout $y \in \mathbb{R}^n$, soit $\nu(y, dx)$ la mesure de probabilité sur $\mathbb{R}^m$ définie par*

$$\nu(y, dx) = \begin{cases} \dfrac{1}{q(y)}\, p(x, y)\, dx & \text{si } q(y) > 0, \\ \delta_0(dx) & \text{si } q(y) = 0. \end{cases}$$

Alors, pour toute fonction $h : \mathbb{R}^m \longrightarrow \mathbb{R}_+$ borélienne,

$$E[h(X)\,|\,Y] = \int \nu(Y, dx)\, h(x).$$

On écrit souvent, de manière un peu abusive, pour tout $y \in \mathbb{R}$,

$$E[h(X)\,|\,Y = y] = \int \nu(y, dx)\, h(x) = \frac{1}{q(y)} \int h(x)\, p(x, y)\, dx$$

et on dit que $\nu(y, dx)$ est la *loi conditionnelle* de X sachant que $Y = y$. La fonction

$$x \longrightarrow \frac{p(x, y)}{q(y)}$$

est appelée *densité conditionnelle* de X sachant que $Y = y$.

Exercice. Sous les hypothèses précédentes, montrer plus généralement que, pour toute fonction borélienne $h : \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R}_+$, on a

$$E[h(X, Y)\,|\,Y] = \int h(x, Y)\, \nu(Y, dx).$$

11.4.3 Conditionnement gaussien

Soient $X, Y_1, \ldots, Y_p$ $p + 1$ variables aléatoires réelles dans $L^2(\Omega, \mathcal{A}, P)$. Comme cela a été vu dans le paragraphe 2.3 ci-dessus, l'espérance conditionnelle

$$E[X\,|\,Y_1, \ldots, Y_p]$$

est la projection orthogonale de X sur l'espace $L^2(\Omega, \sigma(Y_1, \ldots, Y_p), P)$ qui est de dimension infinie sauf dans des cas triviaux. Cette projection orthogonale est aussi la meilleure approximation de X, au sens de la norme L^2, par une v.a. de la forme $\varphi(Y_1, \ldots, Y_p)$.

Par ailleurs, nous avons aussi étudié, dans le Chapitre 8, la meilleure approximation de X par une fonction affine de $Y_1, \ldots, Y_p$, qui est la projection orthogonale de X sur l'espace vectoriel (de dimension finie) engendré par $1, Y_1, \ldots, Y_p$. En général cette dernière projection est très différente de l'espérance conditionnelle $E[X \,|\, Y_1, \ldots, Y_p]$ qui fournit une bien meilleure approximation de X. Nous allons cependant étudier une situation où les deux coïncident, ce qui a l'énorme avantage de ramener les calculs d'espérance conditionnelle à des projections en dimension finie.

Nous avons vu dans le Chapitre 10 qu'une v.a. $Z = (Z_1, \ldots, Z_k)$ à valeurs dans $\mathbb{R}^k$ est un vecteur gaussien centré si toute combinaison linéaire de $Z_1, \ldots, Z_k$ est gaussienne centrée, ce qui équivaut encore à

$$\forall \xi \in \mathbb{R}^k \ , \quad E[\exp(i\xi \cdot Z)] = \exp(-\frac{1}{2}\, {}^t\xi K_Z \xi).$$

C'est par exemple le cas si les composantes $Z_1, \ldots, Z_k$ sont des v.a. gaussiennes indépendantes.

Proposition 11.4.2 *Soit* $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ *un vecteur gaussien centré. Alors les vecteurs* $(X_1, \ldots, X_m)$ *et* $(Y_1, \ldots, Y_n)$ *sont indépendants si et seulement si*

$$\mathrm{cov}(X_i, Y_j) = 0 \ , \qquad \forall i \in \{1, \ldots, m\}, \ j \in \{1, \ldots, n\}. \tag{11.4}$$

Preuve. Il suffit de montrer que, sous la condition (11.4), $(X_1, \ldots, X_m)$ est indépendant de $(Y_1, \ldots, Y_n)$ (l'inverse est toujours vrai). Or, pour $\xi = (\eta_1, \ldots, \eta_m, \zeta_1, \ldots, \zeta_n) \in \mathbb{R}^{n+m}$,

$$E[\exp(i\xi \cdot (X_1, \ldots, X_m, Y_1, \ldots, Y_n))] = \exp(-\frac{1}{2}\, {}^t\xi K_{(X_1,\ldots,X_m,Y_1,\ldots,Y_n)} \xi)$$

et, sous la condition (11.4),

$${}^t\xi K_{(X_1,\ldots,X_m,Y_1,\ldots,Y_n)} \xi = \sum_{j,k=1}^{m} \eta_j \eta_k \mathrm{cov}(X_j, X_k) + \sum_{j,k=1}^{n} \zeta_j \zeta_k \, \mathrm{cov}(Y_j, Y_k).$$

Cela entraîne

$$E[\exp(i\xi \cdot (X_1, \ldots, X_m, Y_1, \ldots, Y_n))] = E[\exp(i \sum_{j=1}^{m} \eta_j X_j)] \, E[\exp(i \sum_{j=1}^{n} \zeta_j Y_j)],$$

soit encore

$$\widehat{P}_{(X_1,\ldots,X_m,Y_1,\ldots,Y_n)}(\eta_1, \ldots, \eta_m, \zeta_1, \ldots, \zeta_n) = \widehat{P}_{(X_1,\ldots,X_m)}(\eta_1, \ldots, \eta_m) \, \widehat{P}_{(Y_1,\ldots,Y_n)}(\zeta_1, \ldots, \zeta_n).$$

En utilisant l'injectivité de la transformée de Fourier, on a donc

$$P_{(X_1,\ldots,X_m,Y_1,\ldots,Y_n)} = P_{(X_1,\ldots,X_m)} \otimes P_{(Y_1,\ldots,Y_n)}$$

ce qui est l'indépendance recherchée. □

Conséquence. Soit $(X_1, \ldots, X_n)$ un vecteur gaussien centré tel que $\mathrm{cov}(X_j, X_k) = 0$ si $j \neq k$. Alors, les v.a. $X_1, \ldots, X_n$ sont indépendantes. En effet, la proposition précédente

entraîne d'abord que X_n est indépendant de $(X_1, \ldots, X_{n-1})$, puis que X_{n-1} est indépendant de $(X_1, \ldots, X_{n-2})$, etc., ce qui permet de conclure.

Plus généralement, si la matrice de covariance de $(X_1, \ldots, X_n)$ est diagonale par blocs de tailles respectives $i_1, \ldots, i_\ell$ (avec $i_1 + \cdots + i_\ell = n$) les sous-vecteurs $(X_1, \ldots, X_{i_1})$, $(X_{i_1+1}, \ldots, X_{i_1+i_2}), \ldots, (X_{i_1+\cdots+i_{\ell-1}+1}, \ldots, X_n)$ sont indépendants.

Théorème 11.4.3 *Soit $(Y_1, \ldots, Y_n, X)$ un vecteur gaussien centré. Alors, $E[X \,|\, Y_1, \ldots, Y_n]$ coïncide avec la projection orthogonale de X sur l'espace vectoriel engendré par $Y_1, \ldots, Y_n$. Il existe donc des réels $\lambda_1, \ldots, \lambda_n$ tels que*

$$E[X \,|\, Y_1, \ldots, Y_n] = \sum_{j=1}^{n} \lambda_j Y_j.$$

De plus, pour toute fonction borélienne $h : \mathbb{R} \longrightarrow \mathbb{R}_+$,

$$E[h(X) \,|\, Y_1, \ldots, Y_n] = \int_{\mathbb{R}} h(x)\, q_{\sum_{j=1}^n \lambda_j Y_j, \sigma^2}(x)\, dx,$$

où

$$\sigma^2 = E[(X - \sum_{j=1}^{n} \lambda_j Y_j)^2]$$

et pour tout $m \in \mathbb{R}$,

$$q_{m,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-m)^2}{2\sigma^2})$$

est la densité de la loi $\mathcal{N}(m, \sigma^2)$.

Remarque. Le cas $\sigma = 0$ se produit si et seulement si $X = \sum_{j=1}^n \lambda_j Y_j$, et alors X est mesurable par rapport à $\sigma(Y_1, \ldots, Y_n)$, de sorte que la deuxième formule du théorème doit s'interpréter comme $E[h(X) \,|\, Y_1, \ldots, Y_n] = h(X)$. Nous écartons ce cas trivial dans la preuve qui suit.

Preuve. Soit $\widehat{X} = \sum_{j=1}^n \lambda_j Y_j$ la projection orthogonale de X sur l'espace vectoriel engendré par $Y_1, \ldots, Y_n$. Alors, pour tout $j \in \{1, \ldots, n\}$,

$$\mathrm{cov}(X - \widehat{X}, Y_j) = E[(X - \widehat{X})Y_j] = 0$$

par définition de la projection orthogonale. Puisque le vecteur $(Y_1, \ldots, Y_n, X - \widehat{X})$ est gaussien centré (toute combinaison linéaire de ses composantes est une combinaison linéaire de $Y_1, \ldots, Y_n, X$), la proposition précédente montre que $X - \widehat{X}$ est indépendant de $Y_1, \ldots, Y_n$. Donc,

$$E[X \,|\, Y_1, \ldots, Y_n] = E[X - \widehat{X} \,|\, Y_1, \ldots, Y_n] + \widehat{X} = E[X - \widehat{X}] + \widehat{X} = \widehat{X}.$$

On a utilisé le fait que $\widehat{X}$ est mesurable par rapport à $\sigma(Y_1, \ldots, Y_n)$, puis l'indépendance de de $X - \widehat{X}$ et de $(Y_1, \ldots, Y_n)$ qui entraîne $E[X - \widehat{X} \,|\, Y_1, \ldots, Y_n] = E[X - \widehat{X}] = 0$.

Pour la dernière assertion, notons $Z = X - \widehat{X}$, de sorte que Z est indépendante de $(Y_1, \ldots, Y_n)$ et suit la loi $\mathcal{N}(0, \sigma^2)$ (Z est gaussienne centrée et par définition $\sigma^2 = E[Z^2]$). On utilise alors le théorème 11.3.4 qui montre que

$$E[h(X) \,|\, Y_1, \ldots, Y_n] = E[h(\sum_{j=1}^{n} \lambda_j Y_j + Z) \,|\, Y_1, \ldots, Y_n] = \int h(\sum_{j=1}^{n} \lambda_j Y_j + z)\, P_Z(dz).$$

En écrivant $P_Z(dz) = q_{0,\sigma^2}(z)dz$ et en faisant un changement de variables évident, on aboutit à la formule de l'énoncé. □

11.5 Probabilités de transition et lois conditionnelles

Les calculs précédents d'espérance conditionnelle peuvent être réénoncés de manière plus agréable à l'aide de la notion de probabilité de transition.

Définition 11.5.1 *Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. On appelle probabilité de transition (ou parfois noyau de transition) de E dans F une application*

$$\nu : E \times \mathcal{F} \longrightarrow [0, 1]$$

qui vérifie les deux propriétés suivantes :

(i) *pour tout $x \in E$, $\nu(x, \cdot)$ est une mesure de probabilité sur $(F, \mathcal{F})$;*

(ii) *pour tout $A \in \mathcal{F}$, l'application $x \longrightarrow \nu(x, A)$ est $\mathcal{E}$-mesurable.*

De manière intuitive, à chaque fois que l'on fixe un point x du premier espace E, la mesure de probabilité $\nu(x, \cdot)$ donne le moyen de choisir de manière aléatoire un point y du deuxième espace F. Dans la théorie des chaînes de Markov, sur laquelle nous reviendrons, on étudie l'évolution au cours du temps d'un phénomène aléatoire dans lequel l'état y à l'instant $n+1$ dépend de l'état x à l'instant n, et d'autres paramètres aléatoires non connus à l'instant n : la loi de l'état à l'instant $n+1$ connaissant l'état à l'instant n est alors fournie par une probabilité de transition $\nu(x, dy)$.

Exemple. Soit λ une mesure positive σ-finie sur $(F, \mathcal{F})$, et soit $f : E \times F \longrightarrow \mathbb{R}_+$ une application mesurable telle que

$$\int_F f(x, y)\, \lambda(dy) = 1 \;, \qquad \forall x \in E.$$

Alors

$$\nu(x, A) = \int_A f(x, y)\, \lambda(dy)$$

définit une probabilité de transition de E dans F. La propriété (ii) de la définition découle en particulier du théorème de Fubini.

Proposition 11.5.1 (i) *Si h est une fonction mesurable positive (ou bornée) sur $(F,\mathcal{F})$, alors*

$$\varphi(x) := \int \nu(x,dy)\,h(y)\ ,\ x\in E$$

est une fonction mesurable positive (ou bornée) sur E.
(ii) *Si λ est une mesure de probabilité sur $(E,\mathcal{E})$, alors*

$$\mu(A) := \int \lambda(dx)\,\nu(x,A)\ ,\ A\in\mathcal{F}$$

est une mesure de probabilité sur $(F,\mathcal{F})$.

La vérification de ces propriétés est facile. Dans (i), on suppose d'abord h étagée, puis on utilise un passage à la limite croissant.

Nous en venons maintenant au lien entre la notion de probabilité de transition et l'espérance conditionnelle.

Définition 11.5.2 *Soient X et Y deux variables aléatoires à valeurs respectivement dans $(E,\mathcal{E})$ et dans $(F,\mathcal{F})$. On appelle loi conditionnelle de Y sachant X toute probabilité de transition ν de E dans F telle que, pour toute fonction h mesurable positive sur $(F,\mathcal{F})$, on ait*

$$E[h(Y)\,|\,X] = \int \nu(X,dy)\,h(y).$$

Remarque. La v.a. $\int \nu(X,dy)\,h(y)$ est obtenue en composant X et l'application $x\longrightarrow \int \nu(x,dy)\,h(y)$, qui est mesurable d'après la proposition précédente. C'est donc bien une fonction de X, comme doit l'être l'espérance conditionnelle $E[h(Y)\,|\,X]$.

Par définition, si ν est une loi conditionnelle de Y sachant X, on a pour tout $A\in\mathcal{F}$,

$$P(Y\in A\,|\,X) = \nu(X,A)\ ,\ \text{p.s.}$$

Il est tentant de remplacer cette égalité de variables aléatoires par l'égalité de nombres réels

$$P(Y\in A\,|\,X=x) = \nu(x,A),$$

pour tout $x\in E$. Bien qu'expliquant l'intuition de la notion de loi conditionnelle, cette dernière égalité n'a en général pas de sens (sauf si X est une v.a. discrète) puisque qu'on aura souvent $P(X=x)=0$ pour tout x, ce qui interdit de définir $P(Y\in A\,|\,X=x)$. La seule formulation correcte est donc la première égalité $P(Y\in A\,|\,X)=\nu(X,A)$.

Discutons maintenant l'unicité de la loi conditionnelle de Y sachant X. Si ν et ν' sont deux lois conditionnelles, on aura, pour tout $A\in\mathcal{F}$,

$$\nu(X,A) = P(Y\in A\,|\,X) = \nu'(X,A)\ ,\ \text{p.s.}$$

ce qui équivaut encore à dire que, pour tout $A\in\mathcal{F}$,

$$\nu(x,A) = \nu'(x,A)\ ,\ P_X(dx)\ \text{p.s.}$$

Supposons que l'espace mesurable $(F, \mathcal{F})$ soit tel qu'une mesure de probabilité sur $(F, \mathcal{F})$ soit caractérisée par ses valeurs sur une famille dénombrable d'ensembles mesurables (c'est le cas pour $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, en considérant les pavés à coordonnées rationnelles). Alors on conclut que

$$\nu(x, \cdot) = \nu'(x, \cdot)\ ,\ \ P_X(dx) \text{ p.s.}$$

Il y a donc unicité en ce sens (et clairement on ne peut pas espérer mieux). Par abus de langage on parlera cependant souvent de *la* loi conditionnelle de Y sachant X.

Considérons maintenant le problème de l'existence de lois conditionnelles.

Théorème 11.5.2 *Supposons que $(E, \mathcal{E})$ et $(F, \mathcal{F})$ soient des espaces métriques complets séparables munis de leur tribu borélienne. Alors il existe toujours une loi conditionnelle de Y sachant X.*

Nous ne démontrerons pas ce théorème qui est un résultat assez difficile de théorie de la mesure. Dans la suite de ce cours, nous n'aurons de toute façon pas besoin du Théorème 11.5.2, car une construction directe permet d'éviter le recours au théorème d'existence. Pour illustrer cela reprenons les exemples traités dans la partie précédente (attention les rôles de X et Y sont intervertis).

(1) Si X est une v.a. discrète, c'est-à-dire si E est dénombrable, alors on peut définir $\nu(x, A)$ par

$$\begin{aligned} &\nu(x, A) = P(Y \in A \,|\, X = x) && \text{si } x \in E' := \{a \in E : P(X = a) > 0) \\ &\nu(x, A) = \delta_{y_0}(A) && \text{si } x \notin E' \end{aligned}$$

où y_0 est un point fixé de F, dont le choix est arbitraire.

(2) Supposons que X et Y sont à valeurs respectivement dans $\mathbb{R}^m$ et dans $\mathbb{R}^n$ et que le couple (X, Y) a pour densité $p(x, y)$, $(x, y) \in \mathbb{R}^m \times \mathbb{R}^n$. La densité de X est alors

$$q(x) = \int_{\mathbb{R}^n} p(x, y)\, dy.$$

La Proposition 11.4.1 montre qu'on peut définir la loi conditionnelle de Y sachant X par

$$\begin{aligned} &\nu(x, A) = \frac{1}{q(x)} \int_A dy\, p(x, y) && \text{si } q(x) > 0 \\ &\nu(x, A) = \delta_0(A) && \text{si } q(x) = 0. \end{aligned}$$

(3) Supposons enfin que $(X_1, \ldots, X_n, Y)$ soit un vecteur gaussien centré, et notons

$$\sum_{j=1}^{n} \lambda_j X_j$$

la projection orthogonale de Y sur l'espace vectoriel engendré par $X_1, \ldots, X_n$. Notons aussi

$$\sigma^2 = E[(Y - \sum_{j=1}^{n} \lambda_j X_j)^2].$$

Le Théorème 11.4.3 montre que la loi conditionnelle de Y sachant $X = (X_1, \ldots, X_n)$ est

$$\nu(x_1, \ldots, x_n; A) = \int_A q_{\sum_{j=1}^n \lambda_j x_j, \sigma^2}(y)\, dy$$

où q_{m,σ^2} est la densité de la loi gaussienne $\mathcal{N}(m, \sigma^2)$. De manière légèrement abusive on dit que conditionnellement à $(X_1, \ldots, X_n)$, Y suit la loi $\mathcal{N}(\sum_{j=1}^n \lambda_j X_j, \sigma^2)$.

Partie III

Processus aléatoires

Chapitre 12

Théorie des martingales à temps discret

12.1 Définitions et exemples

On se place sur un espace de probabilité $(\Omega, \mathcal{F}, P)$. Par définition un *processus aléatoire* est une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires définies sur $(\Omega, \mathcal{F}, P)$. Dans ce chapitre, tous les processus aléatoires seront à valeurs réelles.

Définition 12.1.1 *Une filtration de* $(\Omega, \mathcal{F}, P)$ *est une suite croissante* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *de sous-tribus de* $\mathcal{F}$. *On a donc*

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$$

On dit aussi que $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, P)$ *est un espace de probabilité filtré.*

On interprète souvent le paramètre n comme un temps. La tribu $\mathcal{F}_n$ correspond alors à l'information acquise au temps n.

Exemples. (a) Si $(X_n)_{n\in\mathbb{N}}$ est une suite quelconque de v.a. définies sur $(\Omega, \mathcal{F}, P)$, on définit $\mathcal{F}_n^X$ comme étant la plus petite tribu rendant mesurables les v.a. $X_1, X_2, \ldots, X_n$:

$$\mathcal{F}_n^X = \sigma(X_0, X_1, \ldots, X_n).$$

Alors $(\mathcal{F}_n^X)_{n\in\mathbb{N}}$ est une filtration appelée filtration canonique du processus aléatoire $(X_n)_{n\in\mathbb{N}}$.
(b) Supposons que $\Omega = [0, 1[$, $\mathcal{F}$ est la tribu borélienne sur $[0, 1[$, et P est la mesure de Lebesgue. Posons

$$\mathcal{F}_n = \sigma([\frac{i-1}{2^n}, \frac{i}{2^n}[; i = 1, 2, \ldots, 2^n).$$

Alors $(\mathcal{F}_n)_{n\in\mathbb{N}}$ est une filtration appelée filtration dyadique de $[0, 1[$.

Définition 12.1.2 *Un processus* $(X_n)_{n\in\mathbb{N}}$ *est dit adapté à la filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *si pour tout* $n \in \mathbb{N}$, X_n *est mesurable par rapport à la tribu* $\mathcal{F}_n$.

La filtration canonique est par construction la plus petite filtration qui rende le processus adapté.

Dans toute la suite du chapitre (à l'exception de la partie 6), on fixe un espace de probabilité filtré $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, P)$, dont le choix sera parfois précisé dans les exemples. Les notions qui suivent sont bien entendu relatives à cet espace.

Définition 12.1.3 *Soit $(X_n)_{n\in\mathbb{N}}$ un processus adapté, tel que $E[|X_n|] < \infty$ pour tout $n \in \mathbb{N}$. On dit que le processus $(X_n)_{n\in\mathbb{N}}$ est:*

- *une martingale si, pour tout $n \in \mathbb{N}$,*
$$E[X_{n+1} \mid \mathcal{F}_n] = X_n\,;$$
- *une surmartingale si, pour tout $n \in \mathbb{N}$,*
$$E[X_{n+1} \mid \mathcal{F}_n] \leq X_n\,;$$
- *une sous-martingale si, pour tout $n \in \mathbb{N}$,*
$$E[X_{n+1} \mid \mathcal{F}_n] \geq X_n\,.$$

Une conséquence immédiate de la définition d'une martingale est la propriété apparemment plus forte : pour tous $0 \leq n \leq m$,

$$E[X_m \mid \mathcal{F}_n] = X_n \tag{12.1}$$

Cela est facile à vérifier par récurrence sur la valeur de $m - n$: si $m = n$, la propriété est triviale, si $m = n + 1$, c'est la définition, et si $m - n \geq 2$, une propriété bien connue des espérance conditionnelles donne

$$E[X_m \mid \mathcal{F}_n] = E[E[X_m \mid \mathcal{F}_{m-1}] \mid \mathcal{F}_n] = E[X_{m-1} \mid \mathcal{F}_n].$$

Remarquons que (12.1) entraîne $E[X_m] = E[X_n] = E[X_0]$.

De même, si $(X_n)_{n\in\mathbb{N}})$ est une surmartingale (resp. une sous-martingale), on a pour tous $0 \leq n \leq m$,

$$E[X_m \mid \mathcal{F}_n] \leq X_n \qquad (\text{resp. } E[X_m \mid \mathcal{F}_n] \geq X_n),$$

et donc $E[X_m] \leq E[X_n]$ (resp. $E[X_m] \geq E[X_n]$).

Il est souvent utile d'interpréter une martingale comme un jeu équitable : la variable X_n correspond à l'avoir du joueur à l'instant n, et $\mathcal{F}_n$ est l'information dont dispose le joueur à cet instant (en particulier les résultats des jeux précédents). La propriété de martingale $E[X_{n+1} \mid \mathcal{F}_n] = X_n$ traduit donc le fait que la valeur moyenne de l'avoir à l'instant $n + 1$, lorsqu'on connait le passé jusqu'à l'instant n, est l'avoir à l'instant n (en moyenne le joueur ne perd ni ne gagne). De la même façon, une surmartingale correspond à un jeu défavorable.

Il est évident que si $(X_n)_{n\in\mathbb{N}}$ est une surmartingale, $(-X_n)_{n\in\mathbb{N}}$ est une sous-martingale. Pour cette raison, la plupart des résultats qui suivent et sont énoncés seulement pour des surmartingales ont un analogue immédiat pour des sous-martingales (ou bien inversement).

Exemples. (i) Si $X \in L^1(\Omega, \mathcal{F}, P)$ on pose

$$X_n = E[X \mid \mathcal{F}_n].$$

Alors $(X_n)_{n\in\mathbb{N}}$ est une martingale :

$$E[X_{n+1} \mid \mathcal{F}_n] = E[E[X \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = E[X \mid \mathcal{F}_n] = X_n.$$

Une martingale de ce type est dite *fermée.*
(ii) Si $(X_n)_{n\in\mathbb{N}}$ est une suite décroissante et adaptée de v.a. intégrables, alors $(X_n)_{n\in\mathbb{N}}$ est une surmartingale :

$$E[X_{n+1} \mid \mathcal{F}_n] \leq E[X_n \mid \mathcal{F}_n] = X_n.$$

(iii) *Marche aléatoire sur* $\mathbb{R}$. Soit $x \in \mathbb{R}$ et soit $(Y_n)_{n\geq 1}$ une suite de v.a. réelles indépendantes et de même loi μ, telle que $E[|Y_1|] < \infty$. On pose

$$X_0 = x \quad \text{et} \quad X_n = x + Y_1 + Y_2 + \ldots + Y_n \text{ si } n \geq 1.$$

On définit aussi la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ par

$$\mathcal{F}_0 = \{\varnothing, \Omega\} \quad \text{et} \quad \mathcal{F}_n = \sigma(Y_1, \ldots, Y_n) \text{ si } n \geq 1$$

(c'est en fait la filtration canonique de $(X_n)_{n\in\mathbb{N}}$). Alors $(X_n)_{n\in\mathbb{N}}$ est

- une martingale si $E[Y_1] = 0$;
- une surmartingale si $E[Y_1] \leq 0$;
- une sous-martingale si $E[Y_1] \geq 0$.

En effet, par exemple dans le cas $E[Y_1] = 0$, on a

$$E[X_{n+1} \mid \mathcal{F}_n] = E[X_n + Y_{n+1} \mid \mathcal{F}_n] = X_n + E[Y_{n+1}] = X_n,$$

puisque par construction Y_{n+1} est indépendant de $\mathcal{F}_n$.

Le processus $(X_n)_{n\in\mathbb{N}}$ est appelé marche aléatoire sur $\mathbb{R}$ de loi de saut μ, issue de x.
(iv) Reprenons l'exemple (b) d'espace de probabilité filtré donné ci-dessus. Soit μ une mesure finie sur $[0, 1[$, et rappelons que $P = \lambda$ est la mesure de Lebesgue sur $[0, 1[$. Pour tout entier $n \in \mathbb{N}$, posons

$$f_n = \frac{d\mu}{d\lambda_{|\mathcal{F}_n}}$$

qui désigne la dérivée de Radon-Nikodym de μ par rapport à λ, lorsque μ et λ sont vues comme des mesures sur la tribu $\mathcal{F}_n$ (sur la tribu $\mathcal{F}_n$, toutes les mesures sont absolument continues par rapport à λ). Il est facile de vérifier que

$$f_n(\omega) = \sum_{i=1}^{2^n} \frac{\mu([(i-1)2^{-n}, i2^{-n}[)}{2^{-n}} \mathbf{1}_{[(i-1)2^{-n}, i2^{-n}[}(\omega).$$

Alors $(f_n)_{n\in\mathbb{N}}$ est une martingale : si $A \in \mathcal{F}_n$,

$$E[\mathbf{1}_A f_{n+1}] = \int \mathbf{1}_A(\omega)\, f_{n+1}(\omega)\, d\omega = \mu(A) = \int \mathbf{1}_A(\omega)\, f_n(\omega)\, d\omega = E[\mathbf{1}_A f_n],$$

ce qui suffit pour obtenir $f_n = E[f_{n+1} \mid \mathcal{F}_n]$.

Dans le cas particulier où μ est absolument continue par rapport à λ (sur $\mathcal{F}$), la martingale $(f_n)_{n\in\mathbb{N}}$ est du type considéré en (i) ci-dessus : on vérifie aisément que

$$f_n = E[f \mid \mathcal{F}_n],$$

où f est la dérivée de Radon-Nikodym de μ par rapport à λ.

Deux transformations de martingales.

Proposition 12.1.1 *Soit $\varphi : \mathbb{R} \longrightarrow \mathbb{R}_+$ une fonction convexe, et soit $(X_n)_{n\in\mathbb{N}}$ un processus adapté, tel que $E[\varphi(X_n)] < \infty$ pour tout $n \in \mathbb{N}$.*

(i) *Si (X_n) est une martingale, $(\varphi(X_n))$ est une sous-martingale.*

(ii) *Si (X_n) est une sous-martingale et si φ est croissante, $(\varphi(X_n))$ est une sous-martingale.*

En particulier, si X_n est une martingale, $|X_n|$ est une sous-martingale (ainsi que X_n^2 si $E[X_n^2] < \infty$ pour tout n) et si X_n est une sous-martingale, X_n^+ est encore une sous-martingale.

Preuve. (i) D'après l'inégalité de Jensen pour les espérances conditionnelles,

$$E[\varphi(X_{n+1}) \mid \mathcal{F}_n] \geq \varphi(E[X_{n+1} \mid \mathcal{F}_n]) = \varphi(X_n).$$

(ii) De même, puisque $X_n \leq E[X_{n+1} \mid \mathcal{F}_n]$ et φ est croissante,

$$E[\varphi(X_{n+1}) \mid \mathcal{F}_n] \geq \varphi(E[X_{n+1} \mid \mathcal{F}_n]) \geq \varphi(X_n).$$

Définition 12.1.4 *Une famille $(H_n)_{n\geq 1}$ de v.a. réelles est dite prévisible si, pour tout $n \geq 1$, H_n est bornée et $\mathcal{F}_{n-1}$-mesurable.*

Proposition 12.1.2 *Soit $(X_n)_{n\in\mathbb{N}}$ un processus adapté, et $(H_n)_{n\geq 1}$ une famille prévisible. On pose $(H \cdot X)_0 = 0$ et pour tout entier $n \geq 1$,*

$$(H \cdot X)_n = H_1(X_1 - X_0) + H_2(X_2 - X_1) + \cdots + H_n(X_n - X_{n-1}).$$

Alors,

(i) *Si (X_n) est une martingale, $((H \cdot X)_n)$ est aussi une martingale.*

(ii) *Si (X_n) est une surmartingale (resp. une sous-martingale), et si $H_n \geq 0$ pour tout $n \geq 1$, $((H \cdot X)_n)$ est une surmartingale (resp. une sous-martingale).*

Preuve. (i) Puisque les v.a. H_n sont bornées, il est facile de vérifier que les v.a. $(H \cdot X)_n$ sont intégrables. De plus le processus $((H \cdot X)_n)$ est adapté par construction. Il suffit ensuite de vérifier que, pour tout $n \in \mathbb{N}$,

$$E[(H \cdot X)_{n+1} - (H \cdot X)_n \mid \mathcal{F}_n] = 0.$$

Or $(H \cdot X)_{n+1} - (H \cdot X)_n = H_{n+1}(X_{n+1} - X_n)$ et puisque H_{n+1} est $\mathcal{F}_n$-mesurable, on a

$$E[H_{n+1}(X_{n+1} - X_n) \mid \mathcal{F}_n] = H_{n+1}\, E[X_{n+1} - X_n \mid \mathcal{F}_n] = 0.$$

La preuve de (ii) est analogue. □

Si on interprète (dans le cas d'une martingale) X_n comme l'avoir du joueur à l'instant n, la différence $X_{n+1} - X_n$ s'interprète comme le gain réalisé entre les instants n et $n+1$. On peut imaginer que le joueur à l'instant n modifie sa mise en la multipliant par H_{n+1} (qui doit être $\mathcal{F}_n$-mesurable). Le jeu reste équitable, mais le nouveau gain réalisé entre les instants n et $n+1$ est $H_{n+1}(X_{n+1} - X_n)$. Ceci fournit une explication intuitive de la définition de $(H \cdot X)_n$.

12.2 Temps d'arrêt

Définition 12.2.1 *Une v.a. $T : \Omega \longrightarrow \overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$ est appelée temps d'arrêt (de la filtration $(\mathcal{F}_n)$) si pour tout entier $n \in \mathbb{N}$, on a*

$$\{T = n\} \in \mathcal{F}_n.$$

Il est très facile de voir que cela est équivalent à imposer que pour tout $n \in \mathbb{N}$ on a $\{T \leq n\} \in \mathcal{F}_n$. Dans la suite nous utiliserons indifféremment l'une ou l'autre définition.

Il est important de noter que la valeur $+\infty$ est autorisée. En écrivant

$$\{T = +\infty\} = \Omega \backslash \bigcup_{n \in \mathbb{N}} \{T = n\}$$

on voit que $\{T = +\infty\} \in \mathcal{F}_\infty$, où

$$\mathcal{F}_\infty = \bigvee_{n \in \mathbb{N}} \mathcal{F}_n = \sigma\Big(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n\Big).$$

En revenant à l'interprétation en termes de jeu, les temps d'arrêt sont les instants aléatoires auxquels on peut décider de s'arrêter : le point-clé est que pour décider de s'arrêter à l'instant n, on n'a à sa disposition que l'information acquise à cet instant, c'est-à-dire les événements de $\mathcal{F}_n$. Pour prendre une image tirée de la Bourse, il est impossible de décider de vendre ses actions au moment où elles vont être à leur cours maximum de l'année (cela demanderait de connaître le futur à cet instant !).

Exemples. (i) Si $k \in \mathbb{N}$, le temps constant $T = k$ est évidemment un temps d'arrêt.
(ii) Si $(Y_n)_{n \in \mathbb{N}}$ est un processus adapté, et si A est un borélien de $\mathbb{R}$,

$$T_A := \inf\{n \in \mathbb{N} : Y_n \in A\}$$

est un temps d'arrêt, appelé temps d'entrée dans A. En effet, pour tout entier $n \geq 0$,

$$\{T_A = n\} = \{Y_0 \notin A, Y_1 \notin A, \ldots, Y_{n-1} \notin A, Y_n \in A\} \in \mathcal{F}_n.$$

Remarquons que, dans la définition de T_A, on fait la convention $\inf \varnothing = +\infty$. Cette convention sera constamment utilisée dans la suite.
(iii) En revanche, si on fixe $N > 0$ et on pose

$$L_A := \sup\{n \leq N : Y_n \in A\} \quad (\sup \varnothing = 0 \quad \text{par convention})$$

L_A n'est en général pas un temps d'arrêt. En effet, pour $n \in \{1, \ldots, N-1\}$,

$$\{L_A = n\} = \{Y_n \in A, Y_{n+1} \notin A, \ldots, Y_N \notin A\}$$

n'est a priori pas dans $\mathcal{F}_n$.

Proposition 12.2.1 (i) *Si S et T sont deux temps d'arrêt, $S \vee T$ et $S \wedge T$ sont aussi des temps d'arrêt.*
(ii) *Si $(T_k)_{k \in \mathbb{N}}$ est une suite de temps d'arrêt, alors* $\inf(T_k)$, $\sup(T_k)$, $\limsup(T_k)$ *et* $\liminf(T_k)$ *sont aussi des temps d'arrêt.*

Preuve. (i) On écrit $\{S \wedge T \leq n\} = \{S \leq n\} \cup \{T \leq n\}$ et $\{S \vee T \leq n\} = \{S \leq n\} \cap \{T \leq n\}$.
(ii) De même, $\{\inf(T_k) \leq n\} = \cup\{T_k \leq n\}$ et, par exemple,

$$\{\liminf(T_k) \leq n\} = \bigcap_{m=0}^{\infty} \Big(\bigcup_{k=m}^{\infty} \{T_k \leq n\} \Big).$$

Définition 12.2.2 *Soit T un temps d'arrêt. La tribu du passé jusqu'à l'instant T est*

$$\mathcal{F}_T = \{A \in \mathcal{F} : \forall n \in \mathbb{N}, A \cap \{T = n\} \in \mathcal{F}_n\}.$$

On vérifie aisément que $\mathcal{F}_T$ est une tribu et que $\mathcal{F}_T = \mathcal{F}_n$ si $T = n$.

Proposition 12.2.2 *Soient S et T deux temps d'arrêt avec $S \leq T$. Alors, $\mathcal{F}_S \subset \mathcal{F}_T$.*

Preuve. Soit $A \in \mathcal{F}_S$. Alors, pour tout $n \in \mathbb{N}$,

$$A \cap \{T = n\} = \bigcup_{k=0}^{n} (A \cap \{S = k\}) \cap \{T = n\} \in \mathcal{F}_n.$$

Proposition 12.2.3 *Soit $(Y_n)_{n \in \mathbb{N}}$ un processus adapté, et soit T un temps d'arrêt. Alors la v.a. $\mathbf{1}_{\{T<\infty\}} Y_T$ définie par*

$$\mathbf{1}_{\{T<\infty\}} Y_T(\omega) = \begin{cases} Y_n(\omega) & \text{si } T(\omega) = n \in \mathbb{N} \\ 0 & \text{si } T(\omega) = +\infty \end{cases}$$

est $\mathcal{F}_T$-mesurable.

Preuve. Soit B un borélien de $\mathbb{R}$. Alors, pour tout $n \in \mathbb{N}$,

$$\{\mathbf{1}_{\{T<\infty\}} Y_T \in B\} \cap \{T = n\} = \{Y_n \in B\} \cap \{T = n\} \in \mathcal{F}_n,$$

ce qui montre que $\{\mathbf{1}_{\{T<\infty\}} Y_T \in B\} \in \mathcal{F}_T$. Si $0 \in B$, il suffit d'écrire $\{\mathbf{1}_{\{T<\infty\}} Y_T \in B\} = \{\mathbf{1}_{\{T<\infty\}} Y_T \in B^c\}^c$. □

Lorsque le temps d'arrêt T est fini p.s. on écrira bien sûr simplement Y_T au lieu de $\mathbf{1}_{\{T<\infty\}} Y_T$. En particulier, si T est un temps d'arrêt quelconque, $n \wedge T$ est aussi un temps d'arrêt (lemme 12.2.1) et on déduit de la proposition que $Y_{n\wedge T}$ est $\mathcal{F}_{n\wedge T}$-mesurable donc aussi $\mathcal{F}_n$-mesurable d'après la proposition 12.2.2.

Théorème 12.2.4 (Théorème d'arrêt) *Soit $(X_n)_{n\in\mathbb{N}}$ une martingale (resp. une surmartingale) et soit T un temps d'arrêt. Alors $(X_{n\wedge T})_{n\in\mathbb{N}}$ est aussi une martingale (resp. une surmartingale). En particulier, si le temps d'arrêt T est borné, on a $X_T \in L^1$, et*

$$E[X_T] = E[X_0] \qquad (\textit{resp. } E[X_T] \le E[X_0]).$$

Preuve. Pour tout $n \ge 1$, posons

$$H_n = \mathbf{1}_{\{T \ge n\}} = 1 - \mathbf{1}_{\{T \le n-1\}}.$$

Alors la famille $(H_n)_{n\ge 1}$ est prévisible. Puisque

$$X_{n\wedge T} = X_0 + (H \cdot X)_n$$

la première partie du théorème découle de la proposition 12.1.2. Ensuite, si le temps d'arrêt est borné par N, on a $E[X_T] = E[X_{N\wedge T}] = E[X_0]$ (resp. $\le E[X_0]$ dans le cas d'une surmartingale). □

L'hypothèse que T est borné est nécessaire comme le montre l'exemple simple suivant. Considérons la marche aléatoire $X_n = Y_1 + \cdots + Y_n$ issue de 0 et de loi de saut $P(Y_1 = 1) = P(Y_1 = -1) = 1/2$ (c'est ce qu'on appelle la marche aléatoire simple sur $\mathbb{Z}$, ou encore pile ou face). Alors il découle d'un exemple précédent que $(X_n)_{n\in\mathbb{N}}$ est une martingale. Cependant, si on pose

$$T = \inf\{n \ge 0 : X_n = 1\}$$

on a $T < \infty$ p.s. (cf Proposition 10.2.2 – une autre démonstration sera donnée dans la partie suivante) mais

$$1 = E[X_T] \ne E[X_0] = 0.$$

Bien sûr le temps d'arrêt T n'est pas borné, et il n'y a pas de contradiction avec le théorème.

12.3 Convergence presque sûre des martingales

Nous allons maintenant étudier la convergence presque sûre d'une martingale ou d'une sous-martingale quand $n \to \infty$. Considérons d'abord une suite numérique $\alpha = (\alpha_n)_{n\in\mathbb{N}}$. Pour

tous les réels $a < b$ on introduit deux suites de temps $S_k(\alpha)$ et $T_k(\alpha)$ appartenant à $\overline{\mathbb{N}}$, qui sont définies de la manière suivante : on pose

$$S_1(\alpha) = \inf\{n \geq 0 : \alpha_n \leq a\}$$
$$T_1(\alpha) = \inf\{n \geq S_1(\alpha) : \alpha_n \geq b\}$$

puis, par récurrence,

$$S_{k+1}(\alpha) = \inf\{n \geq T_k(\alpha) : \alpha_n \leq a\}$$
$$T_{k+1}(\alpha) = \inf\{n \geq S_{k+1}(\alpha) : \alpha_n \geq b\}.$$

Bien entendu, on utilise toujours la convention $\inf \varnothing = +\infty$ dans ces définitions. On pose ensuite pour tout entier n,

$$N_n([a,b],\alpha) = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k(\alpha)\leq n\}},$$
$$N_\infty([a,b],\alpha) = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k(\alpha)<\infty\}}.$$

La quantité $N_\infty([a,b],\alpha)$ est le nombre de montées effectuées le long de l'intervalle $[a,b]$ par la suite $(\alpha_n)_{n\in\mathbb{N}}$. Nous utiliserons le lemme simple d'analyse suivant.

Lemme 12.3.1 *La suite $(\alpha_n)_{n\in\mathbb{N}}$ converge dans $\overline{\mathbb{R}}$ ssi pour tout choix des rationnels a et b tels que $a < b$, on a $N_\infty([a,b],\alpha) < \infty$.*

Considérons maintenant un processus adapté $(X_n)_{n\in\mathbb{N}}$. Alors les quantités $S_k(X), T_k(X)$ deviennent des v.a. à valeurs dans $\overline{\mathbb{N}}$, et plus précisément il est facile de vérifier que ce sont des temps d'arrêt. En effet, on a par exemple

$$\{T_k(X) \leq n\} = \bigcup_{0\leq m_1<n_1<\cdots<m_k<n_k\leq n} \{X_{m_1} \leq a, X_{n_1} \geq b, \ldots, X_{m_k} \leq a, X_{n_k} \geq b\},$$

ce qui montre que $\{T_k(X) \leq n\} \in \mathcal{F}_n$.

Il en découle en particulier que $N_n([a,b],X)$ est $\mathcal{F}_n$-mesurable.

Lemme 12.3.2 (Inégalité des nombres de montées de Doob) *Supposons que $(X_n)_{n\in\mathbb{N}}$ est une sous-martingale. Alors, pour tous les réels $a < b$ et pour tout $n \in \mathbb{N}$,*

$$(b-a)\, E[N_n([a,b],X)] \leq E[(X_n - a)^+ - (X_0 - a)^+].$$

Preuve. On pose $Y_n = (X_n - a)^+$. D'après la proposition 12.1.1, $(Y_n)_{n\in\mathbb{N}}$ est encore une sous-martingale.

Pour alléger les notations posons $N_n = N_n([a,b],X)$, et écrivons S_k, T_k au lieu de $S_k(X), T_k(X)$. Définissons alors une famille prévisible $(H_n)_{n\geq 1}$ en posant

$$H_n = \sum_{k=1}^{\infty} \mathbf{1}_{\{S_k<n\leq T_k\}} \leq 1$$

(observer que l'événement $\{S_k < n \leq T_k\} = \{S_k \leq n-1\} \backslash \{T_k \leq n-1\}$ est dans $\mathcal{F}_{n-1}$, parce que S_k et T_k sont des temps d'arrêt). Alors, on vérifie facilement que

$$(H \cdot Y)_n = \sum_{k=1}^{N_n} (Y_{T_k} - Y_{S_k}) + \mathbf{1}_{\{S_{N_n+1} < n\}} (Y_n - Y_{S_{N_n+1}}) \geq \sum_{k=1}^{N_n} (Y_{T_k} - Y_{S_k}) \geq N_n(b-a).$$

La première inégalité est vraie parce que $Y_{S_{N_n+1}} = 0$ sur l'ensemble $\{S_{N_n+1} < \infty\}$, et $Y_n \geq 0$. On a donc en particulier

$$E[(H \cdot Y)_n] \geq (b-a)\, E[N_n].$$

Par ailleurs, si $K_n = 1 - H_n$, $(K_n)_{n \in \mathbb{N}}$ est une famille prévisible positive, et la proposition 12.1.2 montre que $(K \cdot Y)$ est une sous-martingale, d'où $E[(K \cdot Y)_n] \geq E[(K.Y)_0] = 0$.

On observe ensuite que

$$(K \cdot Y)_n + (H \cdot Y)_n = ((K+H) \cdot Y)_n = Y_n - Y_0\,,$$

et donc

$$(b-a)\, E[N_n] \leq E[(H \cdot Y)_n] \leq E[(K \cdot Y)_n + (H \cdot Y)_n] = E[Y_n - Y_0]$$

ce qui est l'inégalité du lemme. □

Théorème 12.3.3 *Soit $(X_n)_{n \in \mathbb{N}}$ une sous-martingale telle que*

$$\sup_{n \in \mathbb{N}} E[(X_n)^+] < \infty. \tag{12.2}$$

Alors la suite X_n converge p.s. quand $n \to \infty$. De plus sa limite X_∞ vérifie $E[|X_\infty|] < \infty$.

Remarque. En écrivant $E[X_n] = E[(X_n)^+] - E[(X_n)^-]$, et en rappelant qu'une sous-martingale vérifie $E[X_n] \geq E[X_0]$, on voit que, pour tout $k \geq 0$,

$$E[(X_k)^-] \leq \Big(\sup_{n \in \mathbb{N}} E[(X_n)^+]\Big) - E[X_0].$$

L'hypothèse (12.2) est donc équivalente à imposer que

$$\sup_{n \in \mathbb{N}} E[|X_n|] < \infty$$

c'est-à-dire que la suite (X_n) est bornée dans L^1.

Preuve. Soient $a, b \in \mathbb{Q}$ tels que $a < b$. D'après le lemme 12.3.2, on a pour tout $n \geq 1$,

$$(b-a)\, E[N_n([a,b], X)] \leq E[(X_n - a)^+] \leq |a| + E[(X_n)^+] \leq |a| + \sup_{k \in \mathbb{N}} E[(X_k)^+].$$

En faisant tendre vers $+\infty$, et en utilisant (12.2), on trouve

$$(b-a)\, E[N_\infty([a,b], X)] < \infty$$

et donc $N_\infty([a,b],X) < \infty$ p.s. Quitte à écarter une réunion dénombrable d'ensembles de probabilité nulle, on obtient ainsi que p.s., pour tous les rationnels $a < b$, $N_\infty([a,b],X) < \infty$. D'après le lemme 12.3.1, cela suffit pour affirmer que p.s. la suite X_n converge dans $\overline{\mathbb{R}}$.

Ensuite, à l'aide du lemme de Fatou, et de la remarque suivant l'énoncé, on a

$$E[|X_\infty|] \leq \liminf_{n\to\infty} E[|X_n|] \leq \sup_{n\in\mathbb{N}} E[|X_n|] < \infty$$

et en particulier $|X_\infty| < \infty$ p.s. $\square$

Corollaire 12.3.4 *Soit $(X_n)_{n\in\mathbb{N}}$ une surmartingale positive. Alors X_n converge p.s. Sa limite X_∞ est dans L^1 et vérifie $X_n \geq E[X_\infty \mid \mathcal{F}_n]$ pour tout $n \in \mathbb{N}$.*

Preuve. On applique le théorème 12.3.3 à $X'_n = -X_n$, en remarquant que l'hypothèse (12.2) est alors trivialement vérifiée. La dernière assertion découle du lemme de Fatou pour les espérances conditionnelles :

$$X_n \geq \liminf_{m\to\infty} E[X_m \mid \mathcal{F}_n] \geq E[\liminf_{m\to\infty} X_m \mid \mathcal{F}_n] = E[X_\infty \mid \mathcal{F}_n].$$

Exemples. (1) Soit $Y_n = 1 + Z_1 + \cdots + Z_n$ une marche aléatoire simple (pile ou face) issue de 1. On a vu que $(Y_n)_{n\in\mathbb{N}}$ est une martingale par rapport à sa filtration canonique. Posons ensuite

$$T = \inf\{n \geq 0 : Y_n = 0\}.$$

Alors T est un temps d'arrêt. Du théorème 12.2.4 on déduit que $X_n = Y_{n\wedge T}$ est une martingale positive, à laquelle on peut appliquer le corollaire. Donc X_n converge p.s. vers X_∞ tel que $X_\infty < \infty$. Puisque sur l'ensemble $\{T = \infty\}$ on a $|X_{n+1} - X_n| = [Y_{n+1} - Y_n| = 1$ pour tout n, cela n'est possible que si $T < \infty$ p.s. Modulo un argument de symétrie évident, cela démontre la propriété qui avait été utilisée dans le dernier exemple de la partie précédente.

Dans ce cas on a $X_\infty = 0$ p.s. et donc l'inégalité $X_n \geq E[X_\infty \mid \mathcal{F}_n] = 0$ n'est pas une égalité, bien que la suite (X_n) soit une martingale.

Cet exemple montre aussi que la convergence du corollaire (ou du théorème précédent) n'a pas forcément lieu dans L^1 : ici $E[X_n] = 1$ pour tout n alors que $E[X_\infty] = 0$.

(2) *Processus de branchement.* Soit μ une mesure de probabilité sur $\mathbb{N}$, telle que

$$m = \sum_{k=1}^{\infty} k\,\mu(k) < \infty.$$

On exclut les cas particuliers où μ est la mesure de Dirac en 1 ou la mesure de Dirac en 0.

Soit ensuite $(\xi_{n,j})_{n,j\in\mathbb{N}}$ une famille de v.a. indépendantes de loi μ. On fixe aussi un entier $\ell \geq 1$ et on définit par récurrence une suite (X_n) de v.a. à valeurs dans $\mathbb{N}$ en posant

$$X_0 = \ell$$
$$X_{n+1} = \sum_{j=1}^{X_n} \xi_{n,j}\ , \qquad \forall n \in \mathbb{N}.$$

La quantité X_n s'interprète comme le nombre d'individus dans une population à la génération n, sachant que le nombre d'enfants de chaque individu suit la loi μ (et les nombres d'enfants des différents individus sont des v.a. indépendantes).

Alors la suite $m^{-n}X_n$ est une martingale relativement à la filtration

$$\mathcal{F}_0 = \{\varnothing, \Omega\}$$
$$\mathcal{F}_n = \sigma(\xi_{k,j} : k < n, j \in \mathbb{N}) \ , \quad \text{si } n \geq 1.$$

En effet, il est facile de voir que le processus (X_n) est adapté (la définition de X_n ne fait intervenir que les $\xi_{k,j}$ pour $k < n$). Ensuite, pour tout $n \geq 0$,

$$E[X_{n+1} \mid \mathcal{F}_n] = E[\sum_{j=1}^{\infty} \mathbf{1}_{\{j \leq X_n\}}\, \xi_{n,j} \mid \mathcal{F}_n] = \sum_{j=1}^{\infty} \mathbf{1}_{\{j \leq X_n\}}\, E[\xi_{n,j} \mid \mathcal{F}_n] = m\, X_n$$

puisque $\xi_{n,j}$ est indépendante de $\mathcal{F}_n$, et donc $E[\xi_{n,j} \mid \mathcal{F}_n] = E[\xi_{n,j}] = m$. En conséquence,

$$E[m^{-(n+1)}X_{n+1} \mid \mathcal{F}_n] = m^{-n}X_n.$$

Cela montre d'abord que les v.a. X_n sont dans L^1 (une récurrence immédiate montre que $E[X_n] = \ell\, m^n$) et ensuite que la suite $m^{-n}X_n$ est une martingale positive.

Distinguons maintenant trois cas :

• $m < 1$. Puisque X_n est à valeurs entières, la convergence de $m^{-n}X_n$ vers une quantité finie n'est possible que si $X_n = 0$ pour tout n assez grand (extinction presque sûre de la population).

• $m = 1$. Dans ce cas X_n est une martingale positive et on a la même conclusion (extinction presque sûre) une fois que l'on a vérifié que

$$P(\exists N \geq 1, p \geq 1 : \forall n \geq N, X_n = p) = 0.$$

Cette dernière assertion est obtenue comme une conséquence facile du lemme de Borel-Cantelli (on utilise le fait que $\mu(1) < 1$).

• $m > 1$. On a

$$m^{-n}X_n \underset{n\to\infty}{\longrightarrow} Z \tag{12.3}$$

et sur l'ensemble $\{Z > 0\}$ on voit que X_n est de l'ordre de m^n quand n est grand. On voudrait alors vérifier que $P(Z > 0) > 0$ (et aussi que $Z > 0$ p.s. sur l'ensemble $\{\liminf X_n > 0\}$ de non-extinction). Remarquons que si la convergence (12.3) a lieu dans L^1 on a $P(Z > 0) > 0$, puisque dans ce cas $E[Z] = \lim m^{-n}E[X_n] = \ell$. On peut montrer (théorème de Kesten-Stygum) que la convergence (12.3) a lieu dans L^1 ssi

$$\sum_{k=1}^{\infty} k\, \log(k)\, \mu(k) < \infty$$

et qu'alors $Z > 0$ p.s. sur l'ensemble de non-extinction. Nous verrons un résultat un peu plus faible dans la partie 4 ci-dessous.

Si $(X_n)_{n\in\mathbb{N}}$ est une martingale bornée dans L^1, on peut lui appliquer le théorème 12.3.3 et obtenir que X_n converge p.s. vers X_∞. Les exemples précédents montrent qu'il n'y a pas nécessairement convergence dans L^1. Le théorème suivant caractérise les martingales pour lesquelles c'est le cas.

Théorème 12.3.5 *Soit $(X_n)_{n\in\mathbb{N}}$ une martingale. Les deux conditions suivantes sont équivalentes:*

(i) *X_n converge vers X_∞ p.s. et dans L^1.*

(ii) *Il existe une v.a. $Z \in L^1(\Omega, \mathcal{F}, P)$ telle que $X_n = E[Z \mid \mathcal{F}_n]$ pour tout $n \in \mathbb{N}$.*

De plus, si ces conditions sont satisfaites, on peut prendre $Z = X_\infty$ dans (ii)*. On dit alors que la martingale $(X_n)_{n\in\mathbb{N}}$ est fermée.*

Preuve. Supposons d'abord (i). En écrivant

$$X_n = E[X_m \mid \mathcal{F}_n]\ , \qquad \forall m > n$$

et en utilisant le fait que l'application $Y \longrightarrow E[Y \mid \mathcal{F}_n]$ est une contraction de L^1 (i.e. $E[|E[Y \mid \mathcal{F}_n]|] \le E[|Y|]$), on trouve en faisant tendre m vers ∞ que $X_n = E[X_\infty \mid \mathcal{F}_n]$.

Inversement, supposons (ii). La suite $(X_n)_{n\in\mathbb{N}}$ est alors bornée dans L^1 et donc converge p.s. d'après le théorème 12.3.3. Pour obtenir la convergence L^1, traitons d'abord le cas où la v.a. Z est bornée par une constante $K < \infty$. Alors, toutes les v.a. X_n sont aussi bornées par K, et le théorème de convergence dominée donne le résultat voulu. Dans le cas général, fixons $\varepsilon > 0$, et choisissons $M > 0$ assez grand pour que

$$E[|Z - Z\,\mathbf{1}_{\{|Z|\le M\}}|] < \varepsilon.$$

Alors, pour tout n,

$$E[|X_n - E[Z\,\mathbf{1}_{\{|Z|\le M\}} \mid \mathcal{F}_n]|] = E[|E[Z - Z\,\mathbf{1}_{\{|Z|\le M\}} \mid \mathcal{F}_n]|] < \varepsilon.$$

D'après le cas borné, la martingale $E[Z\,\mathbf{1}_{\{|Z|\le M\}} \mid \mathcal{F}_n]$ converge dans L^1. Donc on peut choisir n_0 assez grand pour que, pour tous $m, n \ge n_0$,

$$E[|E[Z\,\mathbf{1}_{\{|Z|\le M\}} \mid \mathcal{F}_m] - E[Z\,\mathbf{1}_{\{|Z|\le M\}} \mid \mathcal{F}_n]|] < \varepsilon.$$

En combinant ceci avec la majoration précédente, on trouve que, pour tous $m, n \ge n_0$,

$$E[|X_m - X_n|] < 3\varepsilon.$$

Comme ε était arbitraire, la suite (X_n) est de Cauchy dans L^1. □

Corollaire 12.3.6 *Soit $Z \in L^1(\Omega, \mathcal{F}, P)$. La martingale $X_n = E[Z \mid \mathcal{F}_n]$ converge p.s. et dans L^1 vers $X_\infty = E[Z \mid \mathcal{F}_\infty]$, où $\mathcal{F}_\infty = \bigvee_{n=1}^{\infty} \mathcal{F}_n$.*

Preuve. Compte-tenu du théorème précédent, il reste à montrer que $X_\infty = E[Z \mid \mathcal{F}_\infty]$. Remarquons d'abord que X_∞ est $\mathcal{F}_\infty$-mesurable puisque les v.a. X_n le sont. Ensuite, pour tout $n \in \mathbb{N}$ et $A \in \mathcal{F}_n$, on a

$$E[Z\,\mathbf{1}_A] = E[X_n\,\mathbf{1}_A] = E[X_\infty\,\mathbf{1}_A].$$

Un argument simple de classe monotone (cf Théorème 1.4.1) montre que l'égalité $E[Z\,\mathbf{1}_A] = E[X_\infty\,\mathbf{1}_A]$, vraie pour $A \in \bigcup_{n=1}^{\infty} \mathcal{F}_n$, reste vraie pour $A \in \sigma\Big(\bigcup_{n=1}^{\infty} \mathcal{F}_n\Big) = \mathcal{F}_\infty$. Le résultat recherché découle ensuite de la propriété caractéristique de l'espérance conditionnelle. □

Exemple. Reprenons l'exemple (iv) de la partie 1 : $\Omega = [0,1[$, $\mathcal{F}$ est la tribu borélienne sur $[0,1[$, et $P = \lambda$ est la mesure de Lebesgue. On considère la filtration dyadique

$$\mathcal{F}_n = \sigma([\frac{i-1}{2^n}, \frac{i}{2^n}[; i = 1, 2, \ldots, 2^n).$$

Soit μ une mesure finie sur $[0,1[$, et pour tout entier $n \in \mathbb{N}$,

$$f_n(\omega) = \frac{d\mu}{d\lambda_{|\mathcal{F}_n}}(\omega) = \sum_{i=1}^{2^n} \frac{\mu([(i-1)2^{-n}, i2^{-n}[)}{2^{-n}}\, \mathbf{1}_{[(i-1)2^{-n}, i2^{-n}[}(\omega).$$

On a déjà remarqué que $(f_n)_{n\in\mathbb{N}}$ est une martingale (positive), et on a donc (Corollaire 12.3.4)

$$f_n \xrightarrow[n\to\infty]{\text{p.s.}} f_\infty$$

avec $\int f_\infty d\lambda < \infty$. De plus $f_n \geq E[f_\infty \mid \mathcal{F}_n]$, ce qui montre que, pour tout $A \in \mathcal{F}_n$,

$$\mu(A) = \int f_n \mathbf{1}_A\, d\lambda \geq \int E[f_\infty \mid \mathcal{F}_n] \mathbf{1}_A\, d\lambda = \int f_\infty \mathbf{1}_A\, d\lambda.$$

En utilisant la densité dans L^1 des fonctions continues à support compact (cf Théorème 4.3.1), on vérifie aisément que l'espace des combinaisons linéaires à coefficients positifs de fonctions indicatrices d'intervalles dyadiques est dense dans l'espace $L^1_+([0,1[, \gamma)$ des fonctions γ-intégrables positives, pour toute mesure finie γ sur $[0,1[$. On déduit alors de l'inégalité précédente que, pour toute fonction g mesurable positive bornée sur $[0,1[$, on a

$$\int g\, d\mu \geq \int g f_\infty\, d\lambda.$$

Il en découle que $\nu = \mu - f_\infty \cdot \lambda$ est une mesure positive sur $[0,1[$.

Montrons que ν est étrangère à λ. Pour tout $n \geq 0$, posons

$$h_n = \frac{d\nu}{d\lambda_{|\mathcal{F}_n}} = f_n - E[f_\infty \mid \mathcal{F}_n].$$

Dans cet exemple on a $\mathcal{F}_\infty = \mathcal{F}$ et donc le corollaire 12.3.6 montre que $E[f_\infty \mid \mathcal{F}_n] \longrightarrow f_\infty$ p.s. En conséquence $h_n \longrightarrow 0$ p.s. et donc

$$\lambda\Big(\Big\{x \in [0,1[: \limsup_{n\to\infty} h_n(x) > 0\Big\}\Big) = 0. \tag{12.4}$$

D'autre part, pour tout $\varepsilon > 0$,

$$\nu(\{x \in [0,1[: h_n(x) \leq \varepsilon\}) = \int \mathbf{1}_{\{h_n \leq \varepsilon\}} h_n\, d\lambda \leq \varepsilon,$$

ce qui entraîne

$$\nu\Big(\Big\{x \in [0,1[: \limsup_{n\to\infty} h_n(x) < \varepsilon\Big\}\Big) \le \nu\Big(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\{h_n \le \varepsilon\}\Big) \le \varepsilon.$$

On obtient ainsi

$$\nu\Big(\Big\{x \in [0,1[: \limsup_{n\to\infty} h_n(x) = 0\Big\}\Big) = 0$$

et en comparant avec (12.4) on voit que λ et ν sont portées par des boréliens disjoints.

Finalement l'écriture $\mu = f_\infty \cdot \lambda + \nu$ est la décomposition de Lebesgue de la mesure μ comme somme d'une mesure absolument continue et d'une mesure étrangère à la mesure de Lebesgue. De plus, μ est absolument continue par rapport à λ ssi $\nu = 0$ ce qui équivaut à dire que la martingale (f_n) est fermée.

12.4 La convergence dans L^p pour $p > 1$

Notre but est maintenant d'étudier sous quelles conditions une martingale (X_n) converge dans L^p lorsque $p > 1$. Cela nous amènera à obtenir des estimations importantes pour la probabilité de grandes valeurs du supremum $\sup_{n\in\mathbb{N}} X_n$.

Lemme 12.4.1 *Soit $(X_n)_{n\in\mathbb{N}}$ une sous-martingale, et soient S et T deux temps d'arrêt bornés tels que $S \le T$. Alors*

$$E[X_S] \le E[X_T].$$

Remarque. Le cas $S = 0$ a déjà été vu dans le théorème 12.2.4.
Preuve. On sait déjà que X_S et X_T sont dans L^1. On définit ensuite une famille prévisible en posant, pour tout $n \ge 1$,

$$H_n = \mathbf{1}_{\{S<n\le T\}} = \mathbf{1}_{\{S\le n-1\}} - \mathbf{1}_{\{T\le n-1\}}.$$

Alors, si N est un entier choisi pour que $S \le T \le N$, on a

$$(H\cdot X)_N = X_T - X_S$$

et $E[(H.X)_N] \ge 0$ puisque $H.X$ est une sous-martingale (théorème 12.1.2). □

Théorème 12.4.2 (Inégalité maximale de Doob) *Soit $(X_n)_{n\in\mathbb{N}}$ une sous-martingale. Alors, pour tout $a > 0$ et tout $n \in \mathbb{N}$,*

$$a\,P\Big(\sup_{0\le k\le n} X_k \ge a\Big) \le E\Big[X_n\,\mathbf{1}_{\{\sup_{0\le k\le n} X_k \ge a\}}\Big] \le E[X_n^+].$$

Preuve. Introduisons le temps d'arrêt

$$T = \inf\{n \ge 0 : X_n \ge a\}.$$

Alors, si

$$A = \{ \sup_{0 \le k \le n} X_k \ge a \}$$

on a $A = \{T \le n\}$. Par ailleurs, en appliquant le lemme précédent aux temps d'arrêt $T \wedge n$ et n, on a

$$E[X_{T\wedge n}] \le E[X_n]$$

et d'autre part,

$$X_{T\wedge n} \ge a\,\mathbf{1}_A + X_n\,\mathbf{1}_{A^c}.$$

En combinant ces deux inégalités, on trouve

$$E[X_n] \ge aP(A) + E[X_n\,\mathbf{1}_{A^c}]$$

d'où la première inégalité du théorème. La seconde est immédiate. □

Proposition 12.4.3 *Soit $p > 1$ et soit $(X_n)_{n\in\mathbb{N}}$ une sous-martingale positive. Posons*

$$\widetilde{X}_n = \sup_{0\le k\le n} X_k.$$

Alors, pour tout $n \ge 0$,

$$E[(\widetilde{X}_n)^p] \le (\frac{p}{p-1})^p\, E[(X_n)^p].$$

En conséquence, si $(Y_n)_{n\in\mathbb{N}}$ est une martingale et si

$$Y_n^* = \sup_{0\le k\le n} |Y_k|$$

on a pour tout $n \ge 0$:

$$E[(Y_n^*)^p] \le (\frac{p}{p-1})^p\, E[|Y_n|^p].$$

Preuve. La deuxième partie de la proposition découle de la première appliquée à la sous-martingale $X_n = |Y_n|$. Pour la première partie, on peut supposer $E[(X_n)^p] < \infty$, car sinon il n'y a rien à montrer. Alors, l'inégalité de Jensen pour les espérances conditionelles montre que, pour tout $0 \le k \le n$, on a

$$E[(X_k)^p] \le E[E[X_n \mid \mathcal{F}_k]^p] \le E[E[(X_n)^p \mid \mathcal{F}_n]] = E[(X_n)^p]. \tag{12.5}$$

On a donc aussi $E[(\widetilde{X}_n)^p] < \infty$.

D'après le théorème 12.4.2, on a pour tout $a > 0$

$$a\,P(\widetilde{X}_n \ge a) \le E[X_n\,\mathbf{1}_{\{\widetilde{X}_n \ge a\}}].$$

on multiplie chaque membre de cette inégalité par a^{p-2} et on intègre par rapport à la mesure de Lebesgue da sur $]0,\infty[$. A gauche, il vient

$$\int_0^\infty a^{p-1} P(\widetilde{X}_n \ge a)\, da = E\Big[\int_0^{\widetilde{X}_n} a^{p-1} da\Big] = \frac{1}{p}\, E[(\widetilde{X}_n)^p]$$

en utilisant le théorème de Fubini. De même, à droite on a

$$\begin{aligned}\int_0^\infty a^{p-2}E[X_n\,\mathbf{1}_{\{\widetilde{X}_n\geq a\}}]da &= E\Big[X_n\int_0^{\widetilde{X}_n} a^{p-2}da\Big]\\ &= \frac{1}{p-1}E[X_n(\widetilde{X}_n)^{p-1}]\\ &\leq \frac{1}{p-1}E[(X_n)^p]^{\frac{1}{p}}E[(\widetilde{X}_n)^p]^{\frac{p-1}{p}}.\end{aligned}$$

d'après l'inégalité de Hölder. Il vient donc

$$\frac{1}{p}E[(\widetilde{X}_n)^p] \leq \frac{1}{p-1}E[(X_n)^p]^{\frac{1}{p}}E[(\widetilde{X}_n)^p]^{\frac{p-1}{p}}$$

d'où l'inégalité de la première partie de la proposition (on utilise le fait que $E[(\widetilde{X}_n)^p] < \infty$).
□

Si $(X_n)_{n\in\mathbb{N}}$ est un processus aléatoire, on note

$$X^*_\infty = \sup_{n\in\mathbb{N}} |X_n|.$$

Théorème 12.4.4 *Soit $(X_n)_{n\in\mathbb{N}}$ une martingale. Supposons qu'il existe $p > 1$ tel que*

$$\sup_{n\in\mathbb{N}} E[|X_n|^p] < \infty.$$

Alors, X_n converge p.s. et dans L^p vers une v.a. X_∞ telle que

$$E[|X_\infty|^p] = \sup_{n\in\mathbb{N}} E[|X_n|^p]$$

et on a

$$E[(X^*_\infty)^p] \leq (\frac{p}{p-1})^p\, E[|X_\infty|^p].$$

Preuve. La martingale (X_n) étant bornée dans L^1, on sait déjà que X_n converge p.s. vers X_∞. De plus, la proposition 12.4.3 montre que, pour tout $n \in \mathbb{N}$,

$$E[(X^*_n)^p] \leq (\frac{p}{p-1})^p \sup_{k\in\mathbb{N}} E[|X_k|^p].$$

En passant à la limite croissante qund $n \uparrow \infty$, on a

$$E[(X^*_\infty)^p] \leq (\frac{p}{p-1})^p \sup_{k\in\mathbb{N}} E[|X_k|^p] < \infty$$

et donc $X^*_\infty \in L^p$. Puisque toutes les v.a. $|X_n|$ sont dominées par X^*_∞, le théorème de convergence dominée montre que la suite X_n converge dans L^p vers X_∞. Enfin, puisque la suite $E[|X_n|^p]$ est croissante (cf (12.5)) on a

$$E[|X_\infty|^p] = \lim_{n\to\infty} E[|X_n|^p] = \sup_{n\in\mathbb{N}} E[|X_n|^p].$$

Exemple. Revenons au processus de branchement $(X_n)_{n\in\mathbb{N}}$ introduit dans la partie précédente. On suppose que la loi de reproduction μ satisfait

$$m = \sum_{k=0}^{\infty} k\,\mu(k) \in]1,\infty[$$

et

$$\sum_{k=0}^{\infty} k^2\mu(k) < \infty.$$

On pose aussi $\sigma^2 = \mathrm{var}(\mu) = \sum k^2\mu(k) - m^2$. On a vu que $m^{-n}X_n$ est une martingale. Vérifions que cette martingale est bornée dans L^2. On calcule facilement

$$\begin{aligned} E[X_{n+1}^2 \mid \mathcal{F}_n] &= E\Big[\sum_{j,k=1}^{\infty} \mathbf{1}_{\{j\le X_n,k\le X_n\}}\xi_{n,j}\xi_{n,k} \mid \mathcal{F}_n\Big] \\ &= \sum_{j,k=1}^{\infty} \mathbf{1}_{\{j\le X_n,k\le X_n\}}\,E[\xi_{n,j}\xi_{n,k}] \\ &= \sum_{j,k=1}^{\infty} \mathbf{1}_{\{j\le X_n,k\le X_n\}}\,(m^2+\sigma^2\mathbf{1}_{\{j=k\}}) \\ &= m^2X_n^2+\sigma^2X_n. \end{aligned}$$

On a donc

$$E[X_{n+1}^2] = m^2E[X_n^2] + \ell\sigma^2m^n.$$

En posant $a_n = m^{-2n}E[X_n^2]$, on obtient

$$a_{n+1} = a_n + \ell\sigma^2m^{-n-2}$$

et puisque $m > 1$ la suite (a_n) converge. En conséquence, la martingale $m^{-n}X_n$ est bornée dans L^2. D'après le théorème 12.4.4, cette martingale converge dans L^2 vers Z. En particulier, $E[Z] = E[X_0] = \ell$ et donc $P(Z > 0) > 0$ (il n'est pas très difficile de voir qu'on a en fait $Z > 0$ p.s. sur l'ensemble de non-extinction de la population).

12.5 Uniforme intégrabilité et martingales

Définition 12.5.1 *Une famille $(X_i)_{i\in I}$ de v.a. dans $L^1(\Omega,\mathcal{F},P)$ est dite uniformément intégrable (u.i. en abrégé) si*

$$\lim_{a\to+\infty}\Big(\sup_{i\in I} E[|X_i|\mathbf{1}_{\{|X_i|>a\}}]\Big) = 0.$$

Il est immédiat qu'une famille uniformément intégrable est bornée dans L^1 : il suffit de choisir a assez grand pour que

$$\Big(\sup_{i\in I} E[|X_i|\mathbf{1}_{\{|X_i|>a\}}]\Big) \le 1$$

et d'écrire ensuite $E[|X_i|] \le E[|X_i|\mathbf{1}_{\{|X_i|\le a\}}] + E[|X_i|\mathbf{1}_{\{|X_i|>a\}}] \le a+1$. La réciproque est fausse : une famille bornée dans L^1 n'est pas nécessairement u.i.

Exemples. (1) Une famille réduite à un singleton est u.i. (c'est une conséquence simple du théorème de convergence dominée). Plus généralement, tout sous-ensemble fini de $L^1(\Omega, \mathcal{F}, P)$ est u.i.
(2) Si Z est une v.a. positive dans $L^1(\Omega, \mathcal{F}, P)$, l'ensemble des v.a. X telles que $|X| \le Z$ est u.i. (il suffit en effet de majorer $E[|X|\mathbf{1}_{\{|X|>a\}}] \le E[Z\mathbf{1}_{\{Z>a\}}]$ et d'utiliser l'exemple (1)).
(3) Soit $\Phi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ une fonction telle que $x^{-1}\Phi(x) \longrightarrow +\infty$ quand $x \to +\infty$. Alors, pour tout $C > 0$,

$$\{X \in L^1(\Omega, \mathcal{F}, P) : E[\Phi(|X|)] \le C\}$$

est u.i. En effet, il suffit d'écrire

$$E[|X|\mathbf{1}_{\{|X|>a\}}] \le (\sup_{x>a} \frac{x}{\Phi(x)})\, E[\Phi(|X|)].$$

(4) Si $p > 1$, tout sous-ensemble borné de $L^p(\Omega, \mathcal{F}, P)$ est u.i. C'est le cas particulier de (3) où $\Phi(x) = x^p$.

Le nom "uniformément intégrable" est justifié par la proposition suivante.

Proposition 12.5.1 *Soit $(X_i)_{i\in I}$ une famille bornée dans L^1. Il y a équivalence entre :*

(i) *La famille $(X_i)_{i\in I}$ est u.i.*

(ii) *Pour tout $\varepsilon > 0$, on peut choisir $\delta > 0$ de façon que, pour tout événement $A \in \mathcal{F}$ de probabilité $P(A) < \delta$, on ait*

$$\forall i \in I, \quad E[|X_i|\mathbf{1}_A] < \varepsilon.$$

Preuve. (i)⇒(ii) Soit $\varepsilon > 0$. On peut choisir $a > 0$ assez grand tel que

$$\sup_{i\in I} E[|X_i|\mathbf{1}_{\{|X_i|>a\}}] < \frac{\varepsilon}{2}.$$

Si on pose $\delta = \varepsilon/(2a)$, alors la condition $P(A) < \delta$ entraîne que, pour tout $i \in I$,

$$E[|X_i|\mathbf{1}_A] \le E[|X_i|\mathbf{1}_{A\cap\{|X_i|\le a\}}] + E[|X_i|\mathbf{1}_{\{|X_i|>a\}}] \le aP(A) + \frac{\varepsilon}{2} < \varepsilon.$$

(ii)⇒(i) Soit $C = \sup_{i\in I} E[|X_i|]$. D'après l'inégalité de Markov, pour tout $a > 0$,

$$\forall i \in I, \quad P(|X_i| > a) \le \frac{C}{a}.$$

Soit $\varepsilon > 0$ et choisissons δ pour que la propriété de (ii) soit vérifiée. Alors si a est assez grand pour que $C/a < \delta$, on a

$$\forall i \in I, \quad E[|X_i|\mathbf{1}_{\{|X_i|>a\}}] < \varepsilon$$

d'où l'uniforme intégrabilité. □

Corollaire 12.5.2 *Soit* $X \in L^1(\Omega, \mathcal{F}, P)$. *Alors la famille des espérances conditionnelles* $E[X \mid \mathcal{G}]$ *quand* $\mathcal{G}$ *décrit toutes les sous-tribus de* $\mathcal{F}$ *est u.i.*

Preuve. Soit $\varepsilon > 0$. Puisque le singleton $\{X\}$ est u.i., la proposition précédente permet de choisir $\delta > 0$ tel que, pour tout $A \in \mathcal{F}$ avec $P(A) < \delta$ on ait

$$E[|X|\mathbf{1}_A] \leq \varepsilon.$$

Ensuite, pour tout $a > 0$,

$$P(|E[X \mid \mathcal{G}]| > a) \leq \frac{1}{a}\, E[|E[X \mid \mathcal{G}]|] \leq \frac{E[|X|]}{a}.$$

Donc, si a est suffisamment grand pour que $E[|X|]/a < \delta$, on a en utilisant la propriété caractéristique de l'espérance conditionnelle,

$$E[|E[X \mid \mathcal{G}]|\mathbf{1}_{\{|E[X|\mathcal{G}]|>a\}}] \leq E[E[|X| \mid \mathcal{G}]\mathbf{1}_{\{|E[X|\mathcal{G}]|>a\}}] = E[|X|\mathbf{1}_{\{|E[X|\mathcal{G}]|>a\}}] < \varepsilon$$

ce qui donne l'uniforme intégrabilité recherchée. □

Théorème 12.5.3 *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de v.a. dans* L^1 *qui converge en probabilité vers* X_∞. *Alors il y a équivalence entre :*

(i) *La suite* $(X_n)_{n\in\mathbb{N}}$ *converge dans* L^1 *vers* X_∞.

(ii) *La suite* $(X_n)_{n\in\mathbb{N}}$ *est uniformément intégrable.*

Remarque. Le théorème de convergence dominée affirme qu'une suite $(X_n)_{n\to\infty}$ convergeant p.s. (donc aussi en probabilité) converge dans L^1 à condition que $|X_n| \leq Z$ pour tout n, où $Z \geq 0$ est telle que $E[Z] < \infty$. Cette hypothèse de domination est bien sûr plus forte que l'uniforme intégrabilité (cf exemple (2) ci-dessus), qui donne une condition nécessaire et suffisante pour la convergence dans L^1.

Preuve. (i)⇒(ii) D'abord, la suite $(X_n)_{n\in\mathbb{N}}$ est bornée dans L^1. Ensuite, soit $\varepsilon > 0$. On peut choisir N assez grand tel que, pour tout $n \geq N$,

$$E[|X_n - X_N|] < \frac{\varepsilon}{2}.$$

Puisque l'ensemble fini $\{X_0, X_1, \ldots, X_N\}$ est u.i. on peut choisir $\delta > 0$ assez petit de façon que, pour tout événement A de probabilité $P(A) < \delta$,

$$\forall n \in \{0, 1, \ldots, N\}, \quad E[|X_n|\mathbf{1}_A] < \frac{\varepsilon}{2}.$$

Mais alors, si $n > N$, on a aussi

$$E[|X_n|\mathbf{1}_A] \leq E[|X_N|\mathbf{1}_A] + E[|X_n - X_N|] < \varepsilon.$$

On a vérifié la condition (ii) de la proposition 12.5.1, d'où l'uniforme intégrabilité.

(ii)⇒(i) En utilisant la caractérisation de l'uniforme intégrabilité fournie par la proposition 12.5.1(ii), on voit immédiatement que la famille $(X_n - X_m)_{n,m\in\mathbb{N}}$ est aussi u.i. Donc, si $\varepsilon > 0$ est fixé, on peut choisir a assez grand pour que, pour tous $m, n \in \mathbb{N}$,

$$E[|X_n - X_m| \mathbf{1}_{\{|X_n - X_m| > a\}}] < \varepsilon.$$

Alors, pour tous $m, n \in \mathbb{N}$,

$$\begin{aligned} &E[|X_n - X_m|] \\ &\leq E[|X_n - X_m| \mathbf{1}_{\{|X_n - X_m| \leq \varepsilon\}}] + E[|X_n - X_m| \mathbf{1}_{\{\varepsilon < |X_n - X_m| \leq a\}}] + E[|X_n - X_m| \mathbf{1}_{\{|X_n - X_m| > a\}}] \\ &\leq 2\varepsilon + a\, P(|X_n - X_m| > \varepsilon). \end{aligned}$$

La convergence en probabilité de la suite (X_n) entraîne que

$$P(|X_n - X_m| > \varepsilon) \leq P(|X_n - X_\infty| > \frac{\varepsilon}{2}) + P(|X_m - X_\infty| > \frac{\varepsilon}{2}) \underset{n,m\to\infty}{\longrightarrow} 0.$$

On a ainsi obtenu

$$\limsup_{m,n\to\infty} E[|X_n - X_m|] \leq \varepsilon$$

et puisque ε était arbitraire, cela montre que la suite $(X_n)_{n\in\mathbb{N}}$ est de Cauchy pour la norme L^1. □

Remarque. En conséquence du théorème, si une suite $(X_n)_{n\to\infty}$ converge en probabilité et est bornée dans L^p pour une valeur $p > 1$, alors elle converge dans L^1, et même dans L^q pour tout $q < p$ (appliquer le théorème à $|X_n - X_\infty|^q$).

Application aux martingales. En combinant le théorème précédent avec le théorème 12.3.5, on obtient que les trois conditions suivantes sont équivalentes pour une martingale $(X_n)_{n\in\mathbb{N}}$:

(i) X_n converge vers X_∞ p.s. et dans L^1.

(ii) La suite $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable.

(iii) La martingale est fermée : il existe une v.a. $Z \in L^1(\Omega, \mathcal{F}, P)$ telle que $X_n = E[Z \mid \mathcal{F}_n]$ pour tout $n \in \mathbb{N}$.

Remarquons que (ii) découle aussi de (iii) via le corollaire 12.5.2. En particulier toute martingale uniformément intégrable est fermée, et inversement. Rappelons que dans ce cas on a $X_n = E[X_\infty \mid \mathcal{F}_n]$ pour tout n.

Théorèmes d'arrêt. Si $(X_n)_{n\in\mathbb{N}}$ est un processus adapté qui converge p.s. vers X_∞, on définit X_T pour tout temps d'arrêt T fini ou non en posant

$$X_T = \sum_{n=0}^{\infty} \mathbf{1}_{\{T=n\}} X_n + \mathbf{1}_{\{T=\infty\}} X_\infty.$$

Une extension facile de la proposition 12.2.3 montre que X_T est $\mathcal{F}_T$-mesurable.

Théorème 12.5.4 *Soit $(X_n)_{n\in\mathbb{N}}$ une martingale uniformément intégrable. Alors, pour tout temps d'arrêt T fini ou non,*

$$X_T = E[X_\infty \mid \mathcal{F}_T],$$

et en particulier $E[X_T] = E[X_\infty] = E[X_n]$ pour tout $n \in \mathbb{N}$. Si S et T sont deux temps d'arrêt tels que $S \le T$, on a

$$X_S = E[X_T \mid \mathcal{F}_S].$$

Remarques. (i) Une conséquence du théorème et du corollaire 12.5.2 est que la famille $\{X_T, T \text{ temps d'arrêt}\}$ est u.i.
(ii) Pour une martingale quelconque $(X_n)_{n\in\mathbb{N}}$, on peut appliquer le théorème, pour tout entier $N \ge 0$ fixé, à la martingale arrêtée $(X_{n\wedge N})_{n\in\mathbb{N}}$ qui est u.i. On retrouve ainsi certains des résultats précédents.

Preuve. Vérifions d'abord que $X_T \in L^1$:

$$\begin{aligned}
E[|X_T|] &= \sum_{n=0}^{\infty} E[\mathbf{1}_{\{T=n\}}|X_n|] + E[\mathbf{1}_{\{T=\infty\}}|X_\infty|] \\
&= \sum_{n=0}^{\infty} E[\mathbf{1}_{\{T=n\}}|E[X_\infty \mid \mathcal{F}_n]|] + E[\mathbf{1}_{\{T=\infty\}}|X_\infty|] \\
&\le \sum_{n=0}^{\infty} E[\mathbf{1}_{\{T=n\}}E[|X_\infty| \mid \mathcal{F}_n]] + E[\mathbf{1}_{\{T=\infty\}}|X_\infty|] \\
&= \sum_{n=0}^{\infty} E[\mathbf{1}_{\{T=n\}}|X_\infty|] + E[\mathbf{1}_{\{T=\infty\}}|X_\infty|] \\
&= E[|X_\infty|] < \infty.
\end{aligned}$$

De plus, si $A \in \mathcal{F}_T$,

$$\begin{aligned}
E[\mathbf{1}_A X_T] &= \sum_{n\in\mathbb{N}\cup\{\infty\}} E[\mathbf{1}_{A\cap\{T=n\}}X_T] \\
&= \sum_{n\in\mathbb{N}\cup\{\infty\}} E[\mathbf{1}_{A\cap\{T=n\}}X_n] \\
&= \sum_{n\in\mathbb{N}\cup\{\infty\}} E[\mathbf{1}_{A\cap\{T=n\}}X_\infty] \\
&= E[\mathbf{1}_A X_\infty].
\end{aligned}$$

Dans la première égalité on utilisé le fait que $X_T \in L^1$ pour appliquer le théorème de Fubini et échanger somme et intégrale, et dans la troisième égalité on utilise l'égalité $X_n = E[X_\infty \mid \mathcal{F}_n]$ et la propriété de définition $A \cap \{T = n\} \in \mathcal{F}_n$. Puisque X_T est $\mathcal{F}_T$-mesurable, l'identité précédente suffit à montrer que $X_T = E[X_\infty \mid \mathcal{F}_T]$.

Les autres assertions sont faciles : pour la dernière, l'inclusion $\mathcal{F}_S \subset \mathcal{F}_T$ entraîne que

$$X_S = E[X_\infty \mid \mathcal{F}_S] = E[E[X_\infty \mid \mathcal{F}_T] \mid \mathcal{F}_S] = E[X_T \mid \mathcal{F}_S].$$

Théorème 12.5.5 *Soit $(X_n)_{n\in\mathbb{N}}$ une surmatingale. Supposons que l'une des deux conditions suivantes soit vérifiée :*

(i) *$X_n \geq 0$ pour tout $n \in \mathbb{N}$.*

(ii) *La suite $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable.*

Alors, pour tout temps d'arrêt T, fini ou non, on a $X_T \in L^1$. De plus, si S et T sont deux temps d'arrêt tels que $S \leq T$, on a :

– *dans le cas* (i), $\mathbf{1}_{\{S<\infty\}} X_S \geq E[\mathbf{1}_{\{T<\infty\}} X_T \mid \mathcal{F}_S]$;

– *dans le cas* (ii), $X_S \geq E[X_T \mid \mathcal{F}_S]$.

Preuve. Traitons d'abord le cas (i). On a vu que si le temps d'arrêt T est borné, on a $E[X_T] \leq E[X_0]$ (théorème 12.2.4). Le lemme de Fatou montre alors que pour un temps d'arrêt quelconque,

$$E[X_T] \leq \liminf_{k\to\infty} E[X_{T\wedge k}] \leq E[X_0]$$

et donc $X_T \in L^1$. Soient ensuite S et T deux temps d'arrêt tels que $S \leq T$. Supposons d'abord que S et T sont bornés par l'entier N. Alors le lemme 12.4.1 montre que $E[X_S] \geq E[X_T]$. Plus généralement, pour tout événement $A \in \mathcal{F}_S$, on peut considérer le temps d'arrêt

$$S^A(\omega) = \begin{cases} S(\omega) & \text{si } \omega \in A, \\ N & \text{si } \omega \notin A, \end{cases}$$

et de même le temps d'arrêt T^A (noter que $A \in \mathcal{F}_T$). En écrivant $E[X_{S^A}] \geq E[X_{T^A}]$, on trouve

$$\forall A \in \mathcal{F}_S, \quad E[X_S \mathbf{1}_A] \geq E[X_T \mathbf{1}_A].$$

Revenons au cas général où S et T sont des temps d'arrêt quelconques avec $S \leq T$, et soit $B \in \mathcal{F}_S$. En appliquant ce qui précède aux temps d'arrêt $S \wedge k$, $T \wedge k$ et $A = B \cap \{S \leq k\}$, on trouve

$$E[X_{S\wedge k} \mathbf{1}_{B\cap\{S\leq k\}}] \geq E[X_{T\wedge k} \mathbf{1}_{B\cap\{S\leq k\}}] \geq E[X_{T\wedge k} \mathbf{1}_{B\cap\{T\leq k\}}]$$

puisque $\{S \leq k\} \supset \{T \leq k\}$ et $X_{T\wedge k} \geq 0$. Donc,

$$E[X_S \mathbf{1}_{B\cap\{S\leq k\}}] \geq E[X_T \mathbf{1}_{B\cap\{T\leq k\}}]$$

et en faisant tendre k vers ∞ on trouve par convergence dominée

$$E[X_S \mathbf{1}_{B\cap\{S<\infty\}}] \geq E[X_T \mathbf{1}_{B\cap\{T<\infty\}}].$$

En notant $\widetilde{X}_S = \mathbf{1}_{\{S<\infty\}} X_S$ et $\widetilde{X}_T = \mathbf{1}_{\{T<\infty\}} X_T$ on a donc, pour tout $B \in \mathcal{F}_S$,

$$E[\widetilde{X}_S \mathbf{1}_B] \geq E[\widetilde{X}_T \mathbf{1}_B] = E[E[\widetilde{X}_T \mid \mathcal{F}_S] \mathbf{1}_B].$$

Puisque $\widetilde{X}_S$ est $\mathcal{F}_S$-mesurable, cela entraîne facilement $\widetilde{X}_S \geq E[\widetilde{X}_T \mid \mathcal{F}_S]$.

Traitons maintenant le cas (ii). Puisque $(X_n)_{n\in\mathbb{N}}$ est u.i., $(X_n)_{n\in\mathbb{N}}$ est bornée dans L^1, et donc converge p.s. vers X_∞. La convergence a aussi lieu dans L^1 grâce au théorème 12.5.3. La convergence L^1 permet de passer à la limite $m\to\infty$ dans l'inégalité $X_n \geq E[X_{n+m} \mid \mathcal{F}_n]$, et d'obtenir, pour tout $n\in\mathbb{N}$,

$$X_n \geq E[X_\infty \mid \mathcal{F}_n].$$

Par ailleurs, le corollaire 12.3.6 montre que la martingale fermée $Z_n = E[X_\infty \mid \mathcal{F}_n]$ converge p.s. vers X_∞ (noter que X_∞ est $\mathcal{F}_\infty$-mesurable). Si on pose $Y_n = X_n - Z_n$, $(Y_n)_{n\in\mathbb{N}}$ est une surmartingale positive, telle que $Y_\infty = 0$. Du cas (i) (et du théorème 12.5.4) on déduit d'abord que $X_T = Y_T + Z_T$ est dans L^1, puis que

$$Y_S \geq E[Y_T \mid \mathcal{F}_S]$$

(les fonctions indicatrices sont superflues puisque $Y_\infty = 0$). De plus, en appliquant le théorème 12.5.4 à la martingale u.i. Z_n, on a aussi

$$Z_S = E[Z_T \mid \mathcal{F}_S].$$

En combinant les deux relations obtenues on trouve bien

$$X_S \geq E[X_T \mid \mathcal{F}_S].$$

Exemple. *Ruine du joueur.* Considérons à nouveau une marche aléatoire simple (pile ou face) avec $X_0 = k \geq 0$. Soit $m \geq 1$ un entier tel que $0 \leq k \leq m$. On pose

$$T = \inf\{n \geq 0 : X_n = 0 \text{ ou } X_n = m\}.$$

Il découle d'un exemple traité dans la partie 3 que $T < \infty$ p.s. La martingale $Y_n = X_{n\wedge T}$ est uniformément intégrable, puisque bornée, et on a donc $E[Y_\infty] = E[Y_0] = k$, soit

$$m\, P(X_T = m) = k$$

d'où on déduit facilement que

$$P(X_T = m) = \frac{k}{m} \quad , \quad P(X_T = 0) = 1 - \frac{k}{m}.$$

On peut généraliser au pile ou face "biaisé" : on suppose que $X_n = k + Y_1 + \ldots + Y_n$, où les v.a. Y_i sont indépendantes et de même loi

$$P(Y_1 = 1) = p \quad , \quad P(Y_1 = -1) = 1 - p = q,$$

où $p \in]0,1[\backslash\{\frac{1}{2}\}$. Alors il est facile de vérifier que

$$Z_n = (\frac{q}{p})^{X_n}$$

est une martingale. Si T est défini comme ci-dessus, le fait que la martingale $Z_{n\wedge T}$ converge entraîne que $T < \infty$ p.s. Ensuite en appliquant le théorème d'arrêt à la martingale bornée $Z_{n\wedge T}$, on trouve

$$(\frac{q}{p})^k = E\Big[(\frac{q}{p})^{X_T}\Big] = (\frac{q}{p})^m P(X_T = m) + P(X_T = 0)$$

d'où

$$P(X_T = m) = \frac{(\frac{q}{p})^k - 1}{(\frac{q}{p})^m - 1} \quad , \quad P(X_T = 0) = \frac{(\frac{q}{p})^m - (\frac{q}{p})^k}{(\frac{q}{p})^m - 1}.$$

12.6 Martingales rétrogrades

Une filtration rétrograde est une famille $(\mathcal{F}_n)_{n\in -\mathbb{N}}$ indexée par les entiers *négatifs* de sous-tribus de $\mathcal{F}$, telle que, pour tous $m, n \in -\mathbb{N}$,

$$n \leq m \Rightarrow \mathcal{F}_n \subset \mathcal{F}_m.$$

On notera

$$\mathcal{F}_{-\infty} = \bigcap_{n\in -\mathbb{N}} \mathcal{F}_n$$

qui est encore une sous-tribu de $\mathcal{F}$. Remarquons que, à la différence du cas "direct" étudié précédemment, la tribu $\mathcal{F}_n$ devient de plus en plus petite quand $n \to -\infty$.

Un processus $(X_n)_{n\in -\mathbb{N}}$ indexé par les entiers négatifs est une martingale rétrograde (resp. une surmartingale rétrograde, une sous-martingale rétrograde) relativement à la filtration $(\mathcal{F}_n)_{n\in -\mathbb{N}}$ si X_n est $\mathcal{F}_n$-mesurable et $E[|X_n|] < \infty$ pour tout $n \in -\mathbb{N}$, et si, pour tous $m, n \in -\mathbb{N}$,

$$n \leq m \Rightarrow X_n = E[X_m \mid \mathcal{F}_n] \quad (\text{resp. } X_n \geq E[X_m \mid \mathcal{F}_n],\ X_n \leq E[X_m \mid \mathcal{F}_n]).$$

Théorème 12.6.1 *Soit $(X_n)_{n\in -\mathbb{N}}$ une surmartingale rétrograde. Supposons que*

$$\sup_{n\in -\mathbb{N}} E[|X_n|] < \infty. \tag{12.6}$$

Alors la suite $(X_n)_{n\in -\mathbb{N}}$ est uniformément intégrable et converge p.s. et dans L^1 vers X_∞ quand $n \to -\infty$. De plus, pour tout $n \in -\mathbb{N}$,

$$E[X_n \mid \mathcal{F}_{-\infty}] \leq X_\infty.$$

Remarques. (a) Dans le cas d'une martingale rétrograde, la condition (12.6) est automatiquement satisfaite puisqu'on a $X_n = E[X_0 \mid \mathcal{F}_n]$ et donc $E[|X_n|] \leq E[|X_0|]$ pour tout $n \in -\mathbb{N}$. Pour la même raison, l'uniforme intégrabilité de la suite $(X_n)_{n\in -\mathbb{N}}$, dans le cas d'une martingale, découle du corollaire 12.5.2.
(b) Dans le cas "direct" étudié précédemment, le fait qu'une surmartingale (ou une martingale) soit bornée dans L^1 n'entraîne pas son uniforme intégrabilité : en ce sens le cas rétrograde est très différent du cas direct.

Preuve. Nous commençons par établir la convergence p.s. de la suite $(X_n)_{n\in -\mathbb{N}}$, qui découle à nouveau de l'inégalité sur les nombres de montées de Doob. Fixons un entier $K \geq 1$ et posons pour tout $n \in \{0, 1, \ldots, K\}$

$$Y_n^K = X_{-K+n},$$
$$\mathcal{G}_n^K = \mathcal{F}_{-K+n}.$$

Pour $n > K$, on prend aussi $Y_n^K = X_0$ et $\mathcal{G}_n^K = \mathcal{F}_0$. Alors $(Y_n^K)_{n\in\mathbb{N}}$ est une surmartingale relativement à la filtration $(\mathcal{G}_n^K)_{n\in\mathbb{N}}$. En appliquant le lemme 12.3.2 à la sous-martingale $-Y_n^K$, on trouve pour tous $a < b$,

$$(b-a)\, E[N_K([a,b], -Y^N)] \leq E[(-Y_K^K - a)^+] = E[(-X_0 - a)^+] \leq |a| + E[|X_0|].$$

On vérifie aisément que quand $K \uparrow \infty$, $N_K([a,b], -Y^N)$ croît vers

$$\begin{aligned} N([a,b],-X) \;&:=\; \sup\{k \in \mathbb{N} : \exists m_1 < n_1 < \cdots < m_k < n_k \le 0, \\ &\qquad -X_{m_1} \le a, -X_{n_1} \ge b, \ldots, -X_{m_k} \le a, -X_{n_k} \ge b\} \end{aligned}$$

qui est le nombre total de montées de $(-X_n)_{n\in -\mathbb{N}}$ le long de $[a,b]$. Le théorème de convergence monotone entraîne donc

$$(b-a)\, E[N([a,b],-X)] \le |a| + E[|X_0|] < \infty.$$

On obtient ainsi que $N([a,b],X) < \infty$ pour tous les rationnels $a < b$, p.s. Par une adaptation évidente du lemme 12.3.1 cela entraîne la convergence presque sûre de la suite $(X_n)_{n\in-\mathbb{N}}$ quand $n \to -\infty$, et le lemme de Fatou montre que la limite X_∞ vérifie $E[|X_\infty|] < \infty$.

Montrons maintenant que la suite $(X_n)_{n\in-\mathbb{N}}$ est uniformément intégrable. Soit $\varepsilon > 0$. La suite $(E[X_{-n}])_{n\in\mathbb{N}}$ étant croissante et majorée (grâce à (12.6)) on peut choisir un entier $K \le 0$ assez petit de façon que, pour tout $n \le K$,

$$E[X_n] \le E[X_K] + \frac{\varepsilon}{2}.$$

La famille finie $(X_K, X_{K+1}, \ldots, X_{-1}, X_0)$ étant uniformément intégrable, on peut choisir $a > 0$ assez grand de manière que, pour tout $n \in \{K, K+1, \ldots, -1, 0\}$,

$$E[|X_n|\mathbf{1}_{\{|X_n|>a\}}] < \varepsilon.$$

De plus, on peut choisir $\delta > 0$ assez petit de façon que, pour tout $A \in \mathcal{F}$ tel que $P(A) < \delta$ on ait

$$E[|X_K|\mathbf{1}_A] < \frac{\varepsilon}{2}.$$

Ensuite, si $n < K$,

$$\begin{aligned} E[|X_n|\mathbf{1}_{\{|X_n|>a\}}] &= E[-X_n\mathbf{1}_{\{X_n<-a\}}] + E[X_n\mathbf{1}_{\{X_n>a\}}] \\ &= -E[X_n\mathbf{1}_{\{X_n<-a\}}] + E[X_n] - E[X_n\mathbf{1}_{\{X_n\le a\}}] \\ &\le -E[E[X_K \mid \mathcal{F}_n]\mathbf{1}_{\{X_n<-a\}}] + E[X_K] + \frac{\varepsilon}{2} - E[E[X_K \mid \mathcal{F}_n]\mathbf{1}_{\{X_n\le a\}}] \\ &= -E[X_K\mathbf{1}_{\{X_n<-a\}}] + E[X_K] + \frac{\varepsilon}{2} - E[X_K\mathbf{1}_{\{X_n\le a\}}] \\ &= -E[X_K\mathbf{1}_{\{X_n<-a\}}] + E[X_K\mathbf{1}_{\{X_n>a\}}] + \frac{\varepsilon}{2} \\ &\le E[|X_K|\mathbf{1}_{\{|X_n|>a\}}] + \frac{\varepsilon}{2}. \end{aligned}$$

Dans la première inégalité ci-dessus, on a utilisé la propriété $E[X_n] \le E[X_K] + \varepsilon/2$ et l'inégalité de surmartingale $X_n \ge E[X_K \mid \mathcal{F}_n]$. On observe maintenant que

$$P(|X_n| > a) \le \frac{1}{a}\, E[|X_n|] \le \frac{C}{a}$$

où $C = \sup E[|X_k|]$ est fini par hypothèse. Quitte à choisir a encore plus grand, on peut supposer que $C/a < \delta$, de sorte que

$$E[|X_K|\mathbf{1}_{\{|X_n|>a\}}] < \frac{\varepsilon}{2}$$

et en combinant avec ce qui précède,

$$E[|X_n|\mathbf{1}_{\{|X_n|>a\}}] < \varepsilon,$$

pour tout $n < K$. Comme cette inégalité est aussi vraie pour $n \in \{K, K+1, \ldots, -1, 0\}$, cela termine la preuve de l'uniforme intégrabilité de la suite $(X_n)_{n\in-\mathbb{N}}$.

Le reste de la preuve est facile. L'uniforme intégrabilité et la convergence p.s. entraînent la convergence dans L^1. Ensuite, en écrivant

$$E[X_n\mathbf{1}_A] \leq E[X_m\mathbf{1}_A]$$

pour $m \leq n$ et $A \in \mathcal{F}_{-\infty} \subset \mathcal{F}_m$, et en passant à la limite $m \to -\infty$, on trouve

$$E[X_n\mathbf{1}_A] \leq E[X_\infty\mathbf{1}_A]\ , \quad \forall A \in \mathcal{F}_{-\infty}.$$

On a donc aussi

$$E[E[X_n \mid \mathcal{F}_{-\infty}]\mathbf{1}_A] \leq E[X_\infty\mathbf{1}_A]\ , \quad \forall A \in \mathcal{F}_{-\infty}.$$

et puisque X_∞ est clairement $\mathcal{F}_{-\infty}$-mesurable, cela suffit pour entraîner $E[X_n \mid \mathcal{F}_{-\infty}] \leq X_\infty$.
□

Corollaire 12.6.2 *Soit Z une v.a. dans L^1, et soit $(\mathcal{G}_n)_{n\in\mathbb{N}}$ une suite décroissante de tribus. Alors,*

$$E[Z \mid \mathcal{G}_n] \underset{n\to\infty}{\overset{\text{p.s.},L^1}{\longrightarrow}} E[Z \mid \mathcal{G}_\infty]$$

où

$$\mathcal{G}_\infty = \bigcap_{n\in\mathbb{N}} \mathcal{G}_n.$$

Preuve. Pour tout $n \in \mathbb{N}$, posons $X_{-n} = E[Z \mid \mathcal{G}_n]$ et $\mathcal{F}_{-n} = \mathcal{G}_n$. Alors $(X_n)_{n\in-\mathbb{N}}$ est une martingale relativement à la filtration rétrograde $(\mathcal{F}_n)_{n\in-\mathbb{N}}$. Le théorème assure donc que X_n converge p.s. et dans L^1 quand $n \to -\infty$. De plus, grâce à la dernière assertion du théorème, $X_\infty = E[X_0 \mid \mathcal{F}_{-\infty}] = E[E[Z \mid \mathcal{F}_0] \mid \mathcal{F}_{-\infty}] = E[Z \mid \mathcal{F}_{-\infty}]$. □

Applications. (A) *La loi forte des grands nombres.* Soit $\xi_1, \xi_2, \ldots$ une suite de v.a. réelles indépendantes et de même loi, dans L^1. On pose $S_0 = 0$ et pour tout $n \geq 1$,

$$S_n = \xi_1 + \cdots + \xi_n.$$

On remarque que

$$E[\xi_1 \mid S_n] = \frac{1}{n}\, S_n. \tag{12.7}$$

En effet, on sait qu'il existe une fonction mesurable g telle que $E[\xi_1 \mid S_n] = g(S_n)$. Si $k \in \{1, \ldots, n\}$, le couple (ξ_k, S_n) a même loi que (ξ_1, S_n), de sorte que, pour toute fonction h mesurable bornée,

$$E[\xi_k h(S_n)] = E[\xi_1 h(S_n)] = E[g(S_n)h(S_n)]$$

ce qui montre qu'on a aussi $E[\xi_k \mid S_n] = g(S_n)$. Il en résulte que

$$ng(S_n) = E[\xi_1 + \cdots + \xi_n \mid S_n] = S_n$$

d'où l'identité annoncée (12.7).

On a aussi, pour tout $n \geq 1$,

$$E[\xi_1 \mid S_n, \xi_{n+1}, \xi_{n+2}, \ldots] = \frac{1}{n} S_n. \tag{12.8}$$

Cela découle immédiatement de (12.7) et du lemme suivant, appliqué en prenant $Z = \xi_1$, $\mathcal{H}_1 = \sigma(S_n)$ et $\mathcal{H}_2 = \sigma(\xi_{n+1}, \xi_{n+2}, \ldots)$.

Lemme 12.6.3 *Soit Z une v.a. dans L^1 et soient $\mathcal{H}_1$ et $\mathcal{H}_2$ deux sous-tribus de $\mathcal{F}$. Supposons que $\mathcal{H}_2$ est indépendante de $\sigma(Z) \vee \mathcal{H}_1$. Alors,*

$$E[Z \mid \mathcal{H}_1 \vee \mathcal{H}_2] = E[Z \mid \mathcal{H}_1]$$

La preuve de ce lemme est une application simple du lemme de classe monotone (Théorème 1.4.1) : on voit immédiatement que la propriété $E[\mathbf{1}_A Z] = E[\mathbf{1}_A E[Z \mid \mathcal{H}_1]]$ est vraie pour les ensembles $A \in \mathcal{H}_1 \vee \mathcal{H}_2$ de la forme $A = B \cap C$, avec $B \in \mathcal{H}_1$, $C \in \mathcal{H}_2$, et il en découle que cette propriété est vraie pour tout $A \in \mathcal{H}_1 \vee \mathcal{H}_2$.

On peut maintenant appliquer le corollaire 12.6.2 en prenant $Z = \xi_1$ et pour tout $n \geq 0$,

$$\mathcal{G}_n = \sigma(S_n, \xi_{n+1}, \xi_{n+2}, \ldots),$$

de sorte que $\frac{1}{n} S_n = E[Z \mid \mathcal{G}_n]$ par (12.8). On obtient que la suite $\frac{1}{n} S_n$ converge p.s. et dans L^1. La loi du tout ou rien de Kolmogorov (Théorème 10.2.1) assure que la limite est constante et donc égale à $\lim \frac{1}{n} E[S_n] = E[\xi_1]$.

(B) *La loi du tout ou rien de Hewitt-Savage.* Soit $\xi_1, \xi_2, \ldots$ une suite de v.a. indépendantes et de même loi à valeurs dans un espace mesurable $(E, \mathcal{E})$. L'application $\omega \longrightarrow (\xi_1(\omega), \xi_2(\omega), \ldots)$ définit une v.a. à valeurs dans l'espace produit $E^{\mathbb{N}^*}$, qui est muni de la plus petite tribu rendant mesurables les applications coordonnées $(x_1, x_2, \ldots) \longrightarrow x_i$ pour tout $i \in \mathbb{N}^*$. Une fonction mesurable F définie sur $E^{\mathbb{N}^*}$ est dite symétrique si

$$F(x_1, x_2, x_3, \ldots) = F(x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)}, \ldots)$$

pour toute permutation π de $\mathbb{N}^*$ à support fini.

Théorème 12.6.4 *Si F est une fonction symétrique sur $E^{\mathbb{N}^*}$ la variable aléatoire $F(\xi_1, \xi_2, \ldots)$ est constante p.s.*

Exemple. Supposons les v.a. $\xi_1, \xi_2, \ldots$ à valeurs dans $\mathbb{R}^d$, et considérons la marche aléatoire (en dimension d)

$$X_n = \xi_1 + \cdots + \xi_n.$$

Si B est un borélien de $\mathbb{R}^d$,

$$\mathbf{1}_{\{\mathrm{Card}\{n\geq 1: X_n \in B\}=\infty\}}$$

est une fonction symétrique de $\xi_1, \xi_2, \ldots$. On a donc

$$P(\mathrm{Card}\{n \geq 1 : X_n \in B\} = \infty) = 0 \text{ ou } 1.$$

Preuve. Sans perte de généralité on peut supposer F bornée. On pose

$$\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)\ ,\ \mathcal{G}_n = \sigma(\xi_{n+1}, \xi_{n+2}, \ldots).$$

On note $Y = F(\xi_1, \xi_2, \ldots)$ et on pose pour tout $n \in \mathbb{N}$

$$X_n = E[Y \mid \mathcal{F}_n]\ ,\ Z_n = E[Y \mid \mathcal{G}_n].$$

Alors le corollaire 12.3.6 assure que X_n converge p.s. et dans L^1 vers $E[Y \mid \mathcal{F}_\infty] = Y$, cependant que le corollaire 12.6.2 montre que Z_n converge p.s. et dans L^1 vers $E[Y \mid \mathcal{G}_\infty] = E[Y]$ puisque $\mathcal{G}_\infty$ est grossière (loi du tout ou rien de Kolmogorov). Donc pour tout $\varepsilon > 0$, on peut choisir n assez grand de façon que

$$E[|X_n - Y|] < \varepsilon\ ,\ E[|Z_n - E[Y]|] < \varepsilon. \tag{12.9}$$

D'autre part, il existe une fonction mesurable $g : E^n \longrightarrow \mathbb{R}$ telle que $X_n = g(\xi_1, \ldots, \xi_n)$, et la première borne de (12.9) se traduit par :

$$E[|F(\xi_1, \xi_2, \ldots) - g(\xi_1, \ldots, \xi_n)|] < \varepsilon.$$

Puisque la suite $(\xi_{n+1}, \ldots, \xi_{2n}, \xi_1, \ldots, \xi_n, \xi_{2n+1}, \ldots)$ a même loi que $(\xi_1, \xi_2 \ldots)$, cette borne entraîne aussi que

$$E[|F(\xi_{n+1}, \ldots, \xi_{2n}, \xi_1, \ldots, \xi_n, \xi_{2n+1}, \ldots) - g(\xi_{n+1}, \ldots, \xi_{2n})|] < \varepsilon.$$

Mais $F(\xi_{n+1}, \ldots, \xi_{2n}, \xi_1, \ldots, \xi_n, \xi_{2n+1}, \ldots) = F(\xi_1, \ldots, \xi_n, \xi_{n+1}, \ldots, \xi_{2n}, \xi_{2n+1}, \ldots) = Y$ grâce à la symétrie de F, et on a donc obtenu

$$E[|Y - g(\xi_{n+1}, \ldots, \xi_{2n})|] < \varepsilon. \tag{12.10}$$

En prenant l'espérance conditionnelle par rapport à $\mathcal{G}_n$, on a

$$E[|E[Y \mid \mathcal{G}_n] - E[g(\xi_{n+1}, \ldots, \xi_{2n}) \mid \mathcal{G}_n]|] < \varepsilon,$$

soit

$$E[|Z_n - g(\xi_{n+1}, \ldots, \xi_{2n})|] < \varepsilon. \tag{12.11}$$

En combinant (12.10) et (12.11) avec la deuxième borne de (12.9), on trouve

$$E[|Y - E[Y]|] < 3\varepsilon.$$

Puisque ε était arbitraire on a donc $Y = E[Y]$ p.s. $\square$

Chapitre 13

Chaînes de Markov

13.1 Définition et premières propriétés

Dans tout ce chapitre, E est un espace fini ou dénombrable, qui est muni comme d'habitude de la tribu $\mathcal{P}(E)$. Une matrice stochastique sur E est une famille $(Q(x,y), x, y \in E)$ de nombres réels satisfaisant les deux conditions :

(i) $0 \leq Q(x,y) \leq 1$ pour tous $x, y \in E$;

(ii) pour tout $x \in E$, $\displaystyle\sum_{y\in E} Q(x,y) = 1$.

Cette notion est équivalente à celle de probabilité de transition de E dans E : si on pose

$$\nu(x, A) = \sum_{y\in A} Q(x,y) \ , \qquad x \in E,\ A \subset E,$$

on voit que ν est une probabilité de transition de E dans E (voir le Chapitre 11), et inversement si on part d'une telle probabilité de transition ν, la formule $Q(x,y) = \nu(x, \{y\})$ définit une matrice stochastique sur E.

Pour tout entier $n \geq 1$, on peut définir $Q_n = (Q)^n$: $Q_1 = Q$, et ensuite par récurrence,

$$Q_{n+1}(x,y) = \sum_{z\in E} Q_n(x,z)Q(z,y).$$

On vérifie que Q_n est encore une matrice stochastique sur E. On pose aussi $Q_0(x,y) = \mathbf{1}_{\{x=y\}}$.

Pour toute fonction $f : E \longrightarrow \mathbb{R}_+$, on notera Qf la fonction définie par

$$Qf(x) = \sum_{y\in E} Q(x,y)f(y).$$

Définition 13.1.1 *Soit Q une matrice stochastique sur E, et soit $(X_n)_{n\in\mathbb{N}}$ un processus aléatoire à valeurs dans E. On dit que $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition Q si pour tout entier $n \geq 0$, la loi conditionnelle de X_{n+1} connaissant $(X_0, X_1, \ldots, X_n)$ est $Q(X_n, y)$. De manière équivalente, cela signifie que*

$$P(X_{n+1} = y \mid X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = Q(x_n, y),$$

pour tous $x_0, x_1, \ldots, x_n, y \in E$ tels que $P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) > 0$.

Remarques. (i) En général, la loi conditionnelle de X_{n+1} connaissant $X_0, X_1, \ldots, X_n$ dépend de toutes les variables $X_0, X_1, \ldots, X_n$ et pas seulement de la dernière X_n. Le fait qu'ici cette loi conditionnelle ne dépende que de X_n est ce qu'on appelle la **propriété de Markov** : pour prédire le futur (X_{n+1}) la connaissance du passé ($X_0, X_1, \ldots, X_n$) ne donne pas plus d'information que celle du présent (X_n). Nous verrons plus tard d'autres formes plus précises de la propriété de Markov, qui correspondent à la même idée.

(ii) La fonction $Q(x, \cdot)$ donnant la loi conditionnelle de X_{n+1} sachant que $X_n = x$ ne dépend pas de l'entier n : c'est le caractère homogène de la chaîne de Markov. On pourrait aussi considérer des chaînes de Markov inhomogènes, pour lesquelles le mécanisme de transition entre les instants n et $n+1$ dépend de n.

Proposition 13.1.1 *Un processus $(X_n)_{n\in\mathbb{N}}$ à valeurs dans E est une chaîne de Markov de matrice de transition Q ssi, pour tout $n \geq 0$ et pour tous $x_0, x_1, \ldots, x_n \in E$,*

$$P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = P(X_0 = x_0)Q(x_0, x_1)Q(x_1, x_2)\cdots Q(x_{n-1}, x_n). \quad (13.1)$$

En particulier, on a si $P(X_0 = x_0) > 0$,

$$P(X_n = x_n \mid X_0 = x_0) = Q_n(x_0, x_n).$$

Preuve. Si $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition Q la formule donnée est immédiate par récurrence sur n en écrivant

$$\begin{aligned}&P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n, X_{n+1} = x_{n+1}) = \\ &= P(X_0 = x_0, \ldots, X_n = x_n) \times P(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n).\end{aligned}$$

Inversement, si la formule donnée est vraie, on vérifie immédiatement que

$$\begin{aligned}P(X_{n+1} = y \mid X_0 = x_0, \ldots, X_n = x_n) &= \frac{P(X_0 = x_0)Q(x_0, x_1)\cdots Q(x_{n-1}, x_n)Q(x_n, y)}{P(X_0 = x_0)Q(x_0, x_1)\cdots Q(x_{n-1}, x_n)} \\ &= Q(x_n, y).\end{aligned}$$

La dernière assertion s'obtient en remarquant que

$$Q_n(x_0, x_n) = \sum_{x_1, x_2, \ldots, x_{n-1} \in E} Q(x_0, x_1)Q(x_1, x_2)\cdots Q(x_{n-1}, x_n).$$

Remarque. La formule (13.1) montre que pour une chaîne de Markov $(X_n)_{n\in\mathbb{N}}$, la loi de $(X_0, X_1, \ldots, X_n)$ est complètement déterminée par la connaissance de la loi initiale (la loi de X_0) et de la matrice de transition Q.

La proposition suivante rassemble d'autres propriétés simples des chaînes de Markov. Dans (ii) ci-dessous, on utilise la notation $P(A \mid Z)$ pour désigner l'espérance conditionnelle $E[\mathbf{1}_A \mid Z]$.

Proposition 13.1.2 *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov de matrice de transition Q.*

(i) *Pour tout entier* $n \geq 0$ *et toute fonction mesurable* $f : E \longrightarrow \mathbb{R}_+$,

$$E[f(X_{n+1}) \mid X_0, X_1, \dots, X_n] = E[f(X_{n+1}) \mid X_n] = Qf(X_n).$$

Plus généralement, pour tout sous-ensemble fini $\{i_1, \dots, i_k\}$ *de* $\{0, 1, \dots, n-1\}$, *on a*

$$E[f(X_{n+1}) \mid X_{i_1}, \dots, X_{i_k}, X_n] = Qf(X_n).$$

(ii) *Pour tous les entiers* $n \geq 0, p \geq 1$ *et pour tous* $y_1, \dots, y_p \in E$,

$$P(X_{n+1} = y_1, \dots, X_{n+p} = y_p \mid X_0, \dots, X_n) = Q(X_n, y_1)Q(y_1, y_2) \dots Q(y_{p-1}, y_p),$$

et donc

$$P(X_{n+p} = y_p \mid X_n) = Q_p(X_n, y_p).$$

Si on pose $Y_p = X_{n+p}$ *pour tout* $p \in \mathbb{N}$, *le processus* $(Y_p)_{p\in\mathbb{N}}$ *est encore une chaîne de Markov de matrice de transition* Q.

Preuve. (i) D'après la définition,

$$E[f(X_{n+1}) \mid X_0, X_1, \dots, X_n] = \sum_{y \in E} Q(X_n, y)f(y) = Qf(X_n).$$

Ensuite, si $\{i_1, \dots, i_k\}$ est un sous-ensemble fini de $\{0, 1, \dots, n-1\}$, on a

$$\begin{aligned} E[f(X_{n+1}) \mid X_{i_1}, \dots, X_{i_k}, X_n] &= E[E[f(X_{n+1}) \mid X_0, X_1, \dots, X_n] \mid X_{i_1}, \dots, X_{i_k}, X_n] \\ &= E[Qf(X_n) \mid X_{i_1}, \dots, X_{i_k}, X_n] \\ &= Qf(X_n). \end{aligned}$$

(ii) Il découle immédiatement de (13.1) que

$$P(X_{n+1} = y_1, \dots, X_{n+p} = y_p \mid X_0 = x_0, \dots, X_n = x_n) = Q(x_n, y_1)Q(y_1, y_2) \cdots Q(y_{p-1}, y_p).$$

La formule pour $P(X_{n+p} = y_p \mid X_n)$ en découle en sommant sur les choix possibles de $y_1, \dots, y_{p-1}$. Enfin, pour la dernière assertion, on déduit de ce qui précède que

$$P(Y_0 = y_0, Y_1 = y_1, \dots, Y_p = y_p) = P(X_n = y_0)Q(y_0, y_1)Q(y_1, y_2) \dots Q(y_{p-1}, y_p),$$

et on utilise la caractérisation donnée dans la proposition 13.1.1. □

13.2 Quelques exemples

13.2.1 Variables aléatoires indépendantes

Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a. indépendantes à valeurs dans E, de même loi μ, alors $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition

$$Q(x, y) = \mu(y), \quad \forall x, y \in E.$$

La vérification est immédiate. Ce n'est pas l'exemple le plus intéressant de chaîne de Markov !

13.2.2 Marches aléatoires sur $\mathbb{Z}^d$

Soient $\eta, \xi_1, \xi_2, \ldots, \xi_n, \ldots$ des v.a. indépendantes à valeurs dans $\mathbb{Z}^d$. On suppose que $\xi_1, \xi_2, \ldots$ ont même loi μ et on pose pour tout $n \geq 0$,

$$X_n = \eta + \xi_1 + \xi_2 + \cdots + \xi_n.$$

Alors $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition

$$Q(x,y) = \mu(y-x), \quad \forall x, y \in E.$$

En effet, en remarquant que ξ_{n+1} est indépendante de $(X_0, X_1, \ldots, X_n)$, on a

$$\begin{aligned} &P(X_{n+1} = y \mid X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) \\ &\quad = P(\xi_{n+1} = y - x_n \mid X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) \\ &\quad = P(\xi_{n+1} = y - x_n) \\ &\quad = \mu(y - x_n). \end{aligned}$$

Soit $(e_1, \ldots, e_d)$ la base canonique de $\mathbb{R}^d$. Dans le cas où $\mu(e_i) = \mu(-e_i) = \frac{1}{2d}$ pour tout $i \in \{1, \ldots, d\}$, la chaîne de Markov obtenue est appelée la marche aléatoire simple sur $\mathbb{Z}^d$.

13.2.3 Marche aléatoire simple sur un graphe

Soit $\mathcal{P}_2(E)$ l'ensemble des parties de E à deux éléments, et soit A un sous-ensemble de $\mathcal{P}_2(E)$. Pour tout $x \in E$, on note

$$A_x = \{y \in E : \{x,y\} \in A\}.$$

On suppose que A_x est fini et non vide pour tout $x \in E$. On définit alors une matrice de transition Q sur E en posant pour tous $x, y \in E$,

$$Q(x,y) = \begin{cases} \dfrac{1}{\mathrm{Card}\, A_x} & \text{si } \{x,y\} \in A \\ 0 & \text{sinon.} \end{cases}$$

Une chaîne de Markov de matrice de transition Q est appelée marche aléatoire simple sur le graphe (E, A).

13.2.4 Processus de branchement

Rappelons la définition de ces processus déjà étudiés dans le chapitre précédent. Si μ est une mesure de probabilité sur $\mathbb{N}$, et $\ell \in \mathbb{N}$, on définit par récurrence une suite (X_n) de v.a. à valeurs dans $\mathbb{N}$ en posant

$$\begin{aligned} X_0 &= \ell \\ X_{n+1} &= \sum_{j=1}^{X_n} \xi_{n,j}\,, \qquad \forall n \in \mathbb{N}\,, \end{aligned}$$

où les v.a. $\xi_{n,j},\ n,j \in \mathbb{N}$ sont indépendantes et de loi μ. Alors, $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov sur $E = \mathbb{N}$ de matrice de transition

$$Q(x,y) = \mu^{*x}(y), \quad \forall x,y \in \mathbb{N},$$

où μ^{*x} est la convolution de μ x fois avec elle-même, ou de manière équivalente la loi de la somme de x v.a. indépendantes de loi μ (en particulier μ^{*0} est la mesure de Dirac en 0). En effet, en observant que les v.a. $\xi_{n,j},\ j \in \mathbb{N}$ sont indépendantes de $X_0,\ldots,X_n$, on a

$$\begin{aligned}
&P(X_{n+1} = y \mid X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) \\
&\quad = P(\sum_{j=1}^{x_n} \xi_{n,j} = y \mid X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) \\
&\quad = P(\sum_{j=1}^{x_n} \xi_{n,j} = y) \\
&\quad = \mu^{*x_n}(y).
\end{aligned}$$

13.3 La chaîne de Markov canonique

Nous commençons par un résultat d'existence de chaîne de Markov associée à une matrice de transition donnée.

Proposition 13.3.1 *Soit Q une matrice stochastique sur E. On peut trouver un espace de probabilité $(\Omega', \mathcal{F}', P')$ sur lequel il existe, pour tout $x \in E$, un processus $(X_n^x)_{n\in\mathbb{N}}$ qui est une chaîne de Markov de matrice de transition Q, issue de $X_0^x = x$.*

Preuve. On peut prendre $\Omega' = [0,1[$, muni de la tribu borélienne et de la mesure de Lebesgue. A partir du développement dyadique (propre) d'un réel $\omega \in [0,1[$,

$$\omega = \sum_{n=0}^{\infty} \varepsilon_n(\omega)\, 2^{-n-1}, \quad \varepsilon_n(\omega) \in \{0,1\}$$

on construit une suite $(\varepsilon_n)_{n\in\mathbb{N}}$ de v.a. indépendantes de même loi $P(\varepsilon_n = 1) = P(\varepsilon_n = 0) = 1/2$. Si φ est une injection de $\mathbb{N}\times\mathbb{N}$ dans $\mathbb{N}$, les v.a. $\eta_{i,j} = \varepsilon_{\varphi(i,j)}$, $i,j \in \mathbb{N}$ sont (évidemment) encore indépendantes et de même loi. En posant

$$U_i = \sum_{j=0}^{\infty} \eta_{i,j}\, 2^{-j-1}$$

on obtient une suite $U_0, U_1, U_2, \ldots$ de v.a. indépendantes de loi uniforme sur $[0,1]$ (pour voir que U_i suit la loi uniforme, noter que $\sum_{j=0}^{p} \eta_{i,j}\, 2^{-j-1}$ a même loi que $\sum_{n=0}^{p} \varepsilon_n\, 2^{-n-1}$, pour tout entier p, et faire tendre p vers ∞).

Soit $y_1, y_2, \ldots, y_k, \ldots$ une énumération des éléments de E. Fixons aussi $x \in E$. On pose $X_0^x = x$ puis

$$X_1^x = y_k \quad \text{si} \quad \sum_{1\le j<k} Q(x,y_j) < U_1 \le \sum_{1\le j\le k} Q(x,y_j)$$

de sorte qu'il est clair que $P(X_1^x = y) = Q(x,y)$ pour tout $y \in E$. On continue par récurrence en posant

$$X_{n+1}^x = y_k \quad \text{si} \quad \sum_{1 \le j < k} Q(X_n^x, y_j) < U_{n+1} \le \sum_{1 \le j \le k} Q(X_n^x, y_j).$$

En utilisant l'indépendance des v.a. U_i, on vérifie très facilement que pour tout $k \ge 1$,

$$\begin{aligned}
&P(X_{n+1}^x = y_k \mid X_0^x = x_0, X_1^x = x_1, \dots X_n^x = x_n) \\
&\quad = P\Big(\sum_{1 \le j < k} Q(x_n, y_j) < U_{n+1} \le \sum_{1 \le j \le k} Q(x_n, y_j) \mid X_0^x = x_0, X_1^x = x_1, \dots X_n^x = x_n\Big) \\
&\quad = P\Big(\sum_{1 \le j < k} Q(x_n, y_j) < U_{n+1} \le \sum_{1 \le j \le k} Q(x_n, y_j)\Big) \\
&\quad = Q(x_n, y_k),
\end{aligned}$$

de sorte que $(X_n^x)_{n \in \mathbb{N}}$ est une chaîne de Markov de transition Q. □

Dans la suite, il sera utile de faire un choix canonique de l'espace de probabilité sur lequel sera définie la chaîne de Markov étudiée. On prendra

$$\Omega = E^{\mathbb{N}}.$$

Un élément ω de Ω est donc une suite $\omega = (\omega_0, \omega_1, \omega_2, \dots)$ d'élements de E. Les applications coordonnées X_n, $n \in \mathbb{N}$ sont alors définies par

$$X_n(\omega) = \omega_n.$$

On munit Ω de la plus petite tribu, notée $\mathcal{F}$, qui rende mesurables les applications coordonnées. C'est aussi la tribu engendrée par les "cylindres", c'est-à-dire les ensembles C de la forme

$$C = \{\omega \in \Omega : \omega_0 = x_0, \omega_1 = x_1, \dots, \omega_n = x_n\}$$

où $n \in \mathbb{N}$ et $x_0, x_1, \dots x_n \in E$.

Lemme 13.3.2 *Soit $(G, \mathcal{G})$ un espace mesurable, et soit ψ une application de G dans Ω. Alors ψ est mesurable ssi $X_n \circ \psi$ l'est pour tout $n \in \mathbb{N}$.*

Preuve. Il suffit bien sûr de montrer que si $X_n \circ \psi$ est mesurable pour tout n, alors ψ l'est aussi. Or,

$$\{A \in \mathcal{F} : \psi^{-1}(A) \in \mathcal{G}\}$$

est une tribu sur Ω qui par hypothèse contient tous les ensembles de la forme $X_n^{-1}(y)$, $y \in E$, donc rend mesurables toutes les applications coordonnées X_n. Cette tribu est nécessairement $\mathcal{F}$ tout entière. □

Théorème 13.3.3 *Soit Q une matrice stochastique sur E. Pour tout $x \in E$, il existe une unique probabilité, notée $\mathbb{P}_x$, sur $\Omega = E^{\mathbb{N}}$ telle que sous $\mathbb{P}_x$, le processus des coordonnées $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov de matrice de transition Q, et $\mathbb{P}_x(X_0 = x) = 1$.*

Preuve. Soit $x \in E$. La proposition 13.3.1 permet de construire sur un espace de probabilité $(\Omega', \mathcal{F}', P')$ un processus $(X^x_n)_{n\in\mathbb{N}}$ qui est une chaîne de Markov de transition Q telle que $X^x_0 = x$. On définit alors $\mathbb{P}_x$ comme la mesure image de P' par l'application

$$\begin{array}{lcl} \Omega' & \longrightarrow & \Omega \\ \omega' & \longrightarrow & (X^x_n(\omega'))_{n\in\mathbb{N}}. \end{array}$$

Cette application est mesurable grâce au lemme précédent. On a $\mathbb{P}_x(X_0 = x) = P'(X^x_0 = x) = 1$ et de plus pour tous $x_0, x_1, \dots, x_n \in E$,

$$\begin{aligned} \mathbb{P}_x(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) &= P'(X^x_0 = x_0, X^x_1 = x_1, \dots, X^x_n = x_n) \\ &= P'(X^x_0 = x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \\ &= \mathbb{P}_x(X_0 = x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \end{aligned}$$

ce qui montre que sous $\mathbb{P}_x$ le processus des coordonnées est une chaîne de Markov de transition Q (cf proposition 13.1.1).

Pour l'unicité, on remarque que si $\mathbb{P}'_x$ est une autre mesure de probabilité satisfaisant la propriété du théorème, les mesures $\mathbb{P}_x$ et $\mathbb{P}'_x$ coïncident sur les cylindres. Or les cylindres forment une classe stable par intersection finie et qui engendre la tribu $\mathcal{F}$. Le lemme de classe monotone montre alors que $\mathbb{P}_x = \mathbb{P}'_x$ (cf Corollaire 1.4.2). □

Remarques. (a) De la dernière assertion de la proposition 13.1.1, on déduit que, pour tout $n \geq 0$ et tous $x, y \in E$,

$$\mathbb{P}_x(X_n = y) = Q_n(x, y).$$

(b) Si μ est une mesure de probabilité sur E, on notera

$$\mathbb{P}_\mu = \sum_{x\in E} \mu(x)\, \mathbb{P}_x$$

qui définit une mesure de probabilité sur Ω. En écrivant la formule explicite pour $\mathbb{P}_\mu(X_0 = x_0, \dots, X_n = x_n)$, on vérifie immédiatement que sous $\mathbb{P}_\mu$, $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de transition Q, et X_0 a pour loi μ.

(c) Si $(X'_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition Q et de loi initiale μ, alors pour toute partie mesurable B de $\Omega = E^{\mathbb{N}}$, on a

$$P((X'_n)_{n\in\mathbb{N}} \in B) = \mathbb{P}_\mu(B).$$

En effet cette égalité est vraie lorsque B est un cylindre, et on peut ensuite utiliser le même argument qu'à la fin de la preuve ci-dessus. Cette égalité montre que tous les résultats que nous établirons dans la suite pour la chaîne de Markov canonique (celle fournie par le théorème 13.3.3) se transporteront à une chaîne de Markov quelconque de même matrice de transition.

L'un des avantages importants de la chaîne de Markov canonique est de pouvoir utiliser les *opérateurs de translation.* Pour tout $k \in \mathbb{N}$ on définit l'application $\theta_k : \Omega \longrightarrow \Omega$ en posant

$$\theta_k((\omega_n)_{n\in\mathbb{N}}) = (\omega_{k+n})_{n\in\mathbb{N}}.$$

Le lemme 13.3.2 montre que ces applications sont mesurables.

On note $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$ la filtration canonique sur Ω. On utilise aussi la notation $\mathbb{E}_x$ pour désigner l'espérance sous la probabilité $\mathbb{P}_x$.

Théorème 13.3.4 (Propriété de Markov simple) *Soient F et G deux fonctions mesurables positives sur Ω et soit $n \geq 0$. Supposons que F est $\mathcal{F}_n$-mesurable. Alors, pour tout $x \in E$,*

$$\mathbb{E}_x[F \cdot G \circ \theta_n] = \mathbb{E}_x[F \, \mathbb{E}_{X_n}[G]].$$

De manière équivalente,

$$\mathbb{E}_x[G \circ \theta_n \mid \mathcal{F}_n] = \mathbb{E}_{X_n}[G],$$

ce qu'on peut traduire en disant que la loi conditionnelle de $\theta_n(\omega)$ connaissant $(X_0, X_1, \dots, X_n)$ est $\mathbb{P}_{X_n}$.

Remarque. Cet énoncé se généralise aussitôt au cas où on remplace $\mathbb{E}_x$ par $\mathbb{E}_\mu$ pour n'importe quelle loi initiale μ. Il en sera de même pour l'énoncé suivant.

Preuve. Il suffit de montrer la première assertion, et pour cela de traiter le cas où

$$F = \mathbf{1}_{\{X_0=x_0, X_1=x_1, \dots, X_n=x_n\}}$$

pour $x_0, x_1, \dots, x_n \in E$. Considérons d'abord le cas où G est du même type :

$$G = \mathbf{1}_{\{X_0=y_0, X_1=y_1, \dots, X_p=y_p\}}$$

où $p \geq 0$ et $y_0, \dots, y_p \in E$. Dans ce cas, si $y \in E$,

$$\mathbb{E}_y[G] = \mathbf{1}_{\{y_0=y\}} Q(y_0, y_1) \dots Q(y_{p-1}, y_p)$$

et par ailleurs

$$\begin{aligned} \mathbb{E}_x[F \cdot G \circ \theta_n] &= \mathbb{P}_x(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n, X_n = y_0, X_{n+1} = y_{n+1}, \dots, X_{n+p} = y_p) \\ &= \mathbf{1}_{\{x_0=x\}} Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \, \mathbf{1}_{\{y_0=x_n\}} Q(y_0, y_1) \dots Q(y_{p-1}, y_p) \end{aligned}$$

de sorte qu'on obtient facilement le résultat. Un argument de classe monotone montre ensuite que le résultat reste vrai pour toute fonction $G = 1_A$, $A \in \mathcal{F}$, ce qui permet de conclure. □

Le théorème précédent donne une forme générale de la propriété de Markov (simple) : la loi conditionnelle du futur $\theta_n(\omega)$ connaissant le passé $(X_0, X_1, \dots, X_n)$ ne dépend que du présent X_n. Il sera très important de pouvoir étendre cette propriété au cas où n est remplacé par un temps aléatoire T.

Pour illustrer l'intérêt de cette extension, considérons le problème de savoir si partant d'un point x la chaîne y revient infiniment souvent. Autrement dit, en notant

$$N_x = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=x\}}$$

a-t-on $\mathbb{P}_x(N_x = \infty) = 1$? Il suffit en fait de vérifier que la chaîne revient au moins une fois en x. Si

$$H_x = \inf\{n \geq 1 : X_n = x\}$$

avec la convention habituelle $\inf \varnothing = +\infty$, on a l'équivalence

$$\mathbb{P}_x(N_x = \infty) = 1 \Leftrightarrow \mathbb{P}_x(H_x < \infty) = 1.$$

L'implication $\Rightarrow$ est triviale. Dans l'autre sens, supposons $\mathbb{P}_x(H_x < \infty) = 1$. Modulo l'extension de la propriété de Markov mentionnée ci-dessus, on sait que $\theta_{H_x}(\omega) = (\omega_{H_x(\omega)+n})_{n\in\mathbb{N}}$ a pour loi $\mathbb{P}_x$. Mais alors, en écrivant

$$N_x(\omega) = 1 + N_x(\theta_{H_x}(\omega))$$

on voit que N_x a même loi que $1 + N_x$ sous $\mathbb{P}_x$, ce qui n'est possible que si $\mathbb{N}_x = \infty$, $\mathbb{P}_x$ p.s.

Le théorème qui suit permet de rendre ce raisonnement rigoureux (le résultat obtenu sera repris et détaillé dans la partie suivante).

Théorème 13.3.5 (Propriété de Markov forte) *Soit T un temps d'arrêt de la filtration $(\mathcal{F}_n)$. Soient F et G deux fonctions mesurables positives sur Ω. Supposons que F est $\mathcal{F}_T$-mesurable. Alors, pour tout $x \in E$,*

$$\mathbb{E}_x[\mathbf{1}_{\{T<\infty\}}\, F \cdot G \circ \theta_T] = \mathbb{E}_x[\mathbf{1}_{\{T<\infty\}}\, F\, \mathbb{E}_{X_T}[G]].$$

De manière équivalente,

$$\mathbb{E}_x[\mathbf{1}_{\{T<\infty\}}\, G \circ \theta_T \mid \mathcal{F}_T] = \mathbf{1}_{\{T<\infty\}}\, \mathbb{E}_{X_T}[G].$$

Remarque. La v.a. X_T, définie sur l'ensemble $\mathcal{F}_T$-mesurable $\{T < \infty\}$, est $\mathcal{F}_T$-mesurable (cf Proposition 12.2.3 - dans le chapitre précédent on considère des processus à valeurs réelles, mais l'argument reste le même). La v.a. $\mathbb{E}_{X_T}[G]$, définie aussi sur l'ensemble $\{T < \infty\}$, est la composée des applications $\omega \longrightarrow X_T(\omega)$ et $x \to \mathbb{E}_x[G]$.

Preuve. Pour tout entier $n \geq 0$,

$$\mathbb{E}_x[\mathbf{1}_{\{T=n\}}\, F \cdot G \circ \theta_T] = \mathbb{E}_x[\mathbf{1}_{\{T=n\}}\, F \cdot G \circ \theta_n] = \mathbb{E}_x[\mathbf{1}_{\{T=n\}}\, F\, \mathbb{E}_{X_n}[G]]$$

d'après la propriété de Markov simple (théorème 13.3.4) appliquée en observant que $\mathbf{1}_{\{T=n\}}F$ est $\mathcal{F}_n$-mesurable parce que F est $\mathcal{F}_T$-mesurable (cf définition de la tribu $\mathcal{F}_T$ dans le chapitre précédent). Il suffit ensuite de sommer l'égalité obtenue sur toutes les valeurs de $n \in \mathbb{N}$. □

Corollaire 13.3.6 *Soit T un temps d'arrêt tel que $\mathbb{P}_x(T < \infty) = 1$. Supposons qu'il existe $y \in E$ tel que $\mathbb{P}_x(X_T = y) = 1$. Alors sous $\mathbb{P}_x$, $\theta_T(\omega)$ est indépendant de $\mathcal{F}_T$ et de loi $\mathbb{P}_y$.*

Preuve. Avec les notations du théorème, on a

$$\mathbb{E}_x[F \cdot G(\theta_T(\omega))] = \mathbb{E}_x[F\, \mathbb{E}_{X_T}[G]] = \mathbb{E}_x[F\, \mathbb{E}_y[G]] = \mathbb{E}_x[F]\, \mathbb{E}_y[G]$$

d'où les assertions de l'énoncé. □

13.4 La classification des états

A partir de maintenant, on utilise uniquement (sauf exception, notamment dans les exemples) la chaîne de Markov canonique construite dans le paragraphe précédent. Rappelons la notation : pour $x \in E$,

$$H_x = \inf\{n \geq 1 : X_n = x\}$$
$$N_x = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = x\}}.$$

Proposition 13.4.1 (et définition) *Soit $x \in E$. On a :*

- *ou bien $\mathbb{P}_x(H_x < \infty) = 1$, et alors*

$$N_x = \infty\,, \quad \mathbb{P}_x \ p.s.$$

dans ce cas x est dit récurrent;

- *ou bien $\mathbb{P}_x(H_x < \infty) < 1$, et alors*

$$N_x < \infty\,, \quad \mathbb{P}_x \ p.s.$$

et plus précisément $\mathbb{E}_x[N_x] = 1/\mathbb{P}_x(H_x = \infty) < \infty$; dans ce cas x est dit transitoire.

Preuve. Pour tout entier $k \geq 1$, la propriété de Markov forte montre que

$$\begin{aligned}
\mathbb{P}_x(N_x \geq k+1) &= \mathbb{E}_x[\mathbf{1}_{\{H_x<\infty\}}\,\mathbf{1}_{\{N_x \geq k\}} \circ \theta_{H_x}] \\
&= \mathbb{E}_x[\mathbf{1}_{\{H_x<\infty\}}\,\mathbb{E}_x[\mathbf{1}_{\{N_x \geq k\}}]] \\
&= \mathbb{P}_x(H_x < \infty)\,\mathbb{P}_x(N_x \geq k).
\end{aligned}$$

Puisque $\mathbb{P}_x(N_x \geq 1) = 1$, une récurrence immédiate donne $\mathbb{P}_x(N_x \geq k) = \mathbb{P}_x(H_x < \infty)^{k-1}$. Si $\mathbb{P}_x(H_x < \infty) = 1$ il en découle aussitôt que $\mathbb{P}_x(N_x = \infty) = 1$. Si $\mathbb{P}_x(H_x < \infty) < 1$, on trouve

$$\mathbb{E}_x[N_x] = \sum_{k=1}^{\infty} \mathbb{P}_x(N_x \geq k) = \frac{1}{\mathbb{P}_x(H_x = \infty)} < \infty.$$

Définition 13.4.1 *Le noyau potentiel de la chaîne est la fonction $U : E \times E \longrightarrow [0, \infty]$ définie par*

$$U(x,y) = \mathbb{E}_x[N_y].$$

Proposition 13.4.2 (i) *Pour tous $x, y \in E$,*

$$U(x,y) = \sum_{n=0}^{\infty} Q_n(x,y).$$

(ii) *$U(x,x) = \infty$ si et seulement si x est récurrent.*
(iii) *Pour tous $x, y \in E$, avec $x \neq y$,*

$$U(x,y) = \mathbb{P}_x(H_y < \infty)\,U(y,y).$$

Preuve. La propriété (i) est obtenue en écrivant :

$$U(x,y) = \mathbb{E}_x\Big[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=y\}}\Big] = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y) = \sum_{n=0}^{\infty} Q_n(x,y).$$

La propriété (ii) est une conséquence immédiate de la proposition 13.4.1 et de la définition de U.

Enfin (iii) découle de la propriété de Markov forte :

$$\mathbb{E}_x[N_y] = \mathbb{E}_x[\mathbf{1}_{\{H_y<\infty\}}\, N_y \circ \theta_{H_y}] = \mathbb{E}_x[\mathbf{1}_{\{H_y<\infty\}}\, \mathbb{E}_y[N_y]] = \mathbb{P}_x(H_y < \infty)\, U(y,y).$$

Exemple. Considérons la chaîne de Markov sur $\mathbb{Z}^d$ de matrice de transition

$$Q((x_1,\ldots,x_d),(y_1,\ldots,y_d)) = \frac{1}{2^d}\prod_{i=1}^{d} \mathbf{1}_{\{|y_i-x_i|=1\}}$$

(c'est un cas particulier de marche aléatoire sur $\mathbb{Z}^d$). Cette chaîne de Markov issue de 0 a même loi que $(Y_n^1,\ldots,Y_n^d)_{n\in\mathbb{N}}$, où les processus $Y^1,\ldots,Y^d$ sont des copies indépendantes de la marche aléatoire simple (pile ou face) sur $\mathbb{Z}$, issue de 0. En conséquence,

$$Q_n(0,0) = P(Y_n^1 = 0,\ldots,Y_n^d = 0) = P(Y_n^1 = 0)^d.$$

Or $P(Y_n^1 = 0) = 0$ si n est impair, et si $n = 2k$ est pair, un argument de dénombrement simple montre que

$$P(Y_{2k}^1 = 0) = 2^{-2k}\, C_{2k}^k.$$

En conséquence,

$$U(0,0) = \sum_{k=0}^{\infty} Q_{2k}(0,0) = \sum_{k=0}^{\infty} (2^{-2k} C_{2k}^k)^d.$$

La formule de Stirling montre que

$$2^{-2k} C_{2k}^k \underset{k\to\infty}{\sim} \frac{(\frac{2k}{e})^{2k}\sqrt{4\pi k}}{2^{2k}((\frac{k}{e})^k\sqrt{2\pi k})^2} \underset{k\to\infty}{\sim} \sqrt{\frac{1}{\pi k}}.$$

Donc 0 est récurrent si $d = 1$ ou 2, et transitoire si $d \geq 3$.

On note R l'ensemble des états (points) récurrents.

Lemme 13.4.3 *Soit $x \in R$ et soit y un autre point de E tel que $U(x,y) > 0$. Alors $y \in R$ et $\mathbb{P}_y(H_x < \infty) = 1$, donc en particulier $U(y,x) > 0$.*

Preuve. Montrons d'abord que $\mathbb{P}_y(H_x < \infty) = 1$. Pour cela on écrit

$$\begin{aligned}
0 = \mathbb{P}_x(N_x < \infty) &\geq \mathbb{P}_x(H_y < \infty, H_x \circ \theta_{H_y} = \infty) \\
&= \mathbb{E}_x[\mathbf{1}_{\{H_y<\infty\}}\, \mathbf{1}_{\{H_x=\infty\}} \circ \theta_{H_y}] \\
&= \mathbb{E}_x[\mathbf{1}_{\{H_y<\infty\}}\, \mathbb{P}_y(H_x = \infty)] \\
&= \mathbb{P}_x(H_y < \infty)\, \mathbb{P}_y(H_x = \infty).
\end{aligned}$$

L'hypothèse $U(x,y)>0$ entraîne $\mathbb{P}_x(H_y<\infty)>0$. On conclut que $\mathbb{P}_y(H_x=\infty)=0$.

Ensuite, on peut trouver des entiers $n_1,n_2\geq 1$ tels que $Q_{n_1}(x,y)>0$, et $Q_{n_2}(y,x)>0$. Pour tout entier $p\geq 0$, on a alors

$$Q_{n_2+p+n_1}(y,y)\geq Q_{n_2}(y,x)Q_p(x,x)Q_{n_1}(x,y)$$

et donc

$$U(y,y)\geq \sum_{p=0}^{\infty} Q_{n_2+p+n_1}(y,y)\geq Q_{n_2}(y,x)\Big(\sum_{p=0}^{\infty} Q_p(x,x)\Big)Q_{n_1}(x,y)=\infty$$

puisque $x\in R$ entraîne $\sum_{p=0}^{\infty} Q_p(x,x)=U(x,x)=\infty$. □

En conséquence du lemme, si $x\in R$ et $y\in E\backslash R$ on a $U(x,y)=0$: on ne peut pas passer d'un point récurrent à un point transitoire. Cette propriété joue un rôle important dans le théorème suivant.

Théorème 13.4.4 (Classification des états) *Il existe une partition de R*

$$R=\bigcup_{i\in I} R_i$$

telle qu'on ait les propriétés suivantes :

- *si $x\in R$, et si $i\in I$ est tel que $x\in R_i$, on a $\mathbb{P}_x$ p.s.*
  - — $N_y=+\infty\,,\qquad \forall y\in R_i$;
  - — $N_y=0\,,\qquad \forall y\in E\backslash R_i$;
- *si $x\in E\backslash R$ et $T=\inf\{n\geq 0 : X_n\in R\}$, on a $\mathbb{P}_x$ p.s.*
  - *— ou bien $T=\infty$ et $N_y<\infty$, $\forall y\in E$;*
  - *— ou bien $T<\infty$ et il existe un indice (aléatoire) $j\in I$ tel que : $\forall n\geq T,\ X_n\in R_j$.*

Preuve. Pour $x,y\in R$, notons $x\sim y$ si $U(x,y)>0$. Il découle du lemme précédent qu'on ainsi défini une relation d'équivalence sur R (pour la transitivité, on observe que $Q_n(x,y)>0$ et $Q_m(y,z)>0$ entraînent $Q_{n+m}(x,z)>0$. La partition du théorème correspond alors aux classes d'équivalence pour cette relation d'équivalence, qu'on appelle aussi les classes de récurrence de la chaîne de Markov.

Soit $i\in I$ et $x\in R_i$. On a $U(x,y)=0$ pour tout $y\in E\backslash R_i$ (dans le cas $y\in E\backslash R$ on utilise le lemme) et donc $N_y=0$, $\mathbb{P}_x$ p.s. pour tout $y\in E\backslash R_i$. En revanche, si $y\in R_i$, on a $\mathbb{P}_x(H_y<\infty)=1$ d'après le lemme, et la propriété de Markov forte montre que

$$\mathbb{P}_x(N_y=\infty)=\mathbb{E}_x(\mathbf{1}_{\{H_y<\infty\}}\,\mathbf{1}_{\{N_y=\infty\}}\circ\theta_{H_y}]=\mathbb{P}_x(H_y<\infty)\,\mathbb{P}_y(N_y=\infty)=1.$$

Si $x\in E\backslash R$ et $T=\infty$, alors on déduit facilement de la propriété de Markov forte que $N_y<\infty$ pour tout $y\in E\backslash R$. Si $T<\infty$, notons j l'indice (aléatoire) tel que $X_T\in R_j$. En appliquant la propriété de Markov forte en T, et la première partie de l'énoncé, on obtient aisément que $X_n\in R_j$ pour tout $n\geq T$. □

Définition 13.4.2 *La chaîne est dite irréductible si $U(x,y) > 0$ pour tous $x, y \in E$.*

Corollaire 13.4.5 *Si la chaîne est irréductible :*

- *ou bien tous les états sont récurrents, il existe une seule classe de récurrence et on a pour tout $x \in E$,*
$$\mathbb{P}_x(N_y = \infty,\ \forall y \in E) = 1.$$
- *ou bien tous les états sont transitoires et alors, pour tout $x \in E$,*
$$\mathbb{P}_x(N_y < \infty,\ \forall y \in E) = 1.$$

Lorsque E est fini, seul le premier cas peut se produire.

Preuve. S'il existe un état récurrent, le lemme 13.4.3 montre aussitôt que tous les états sont récurrents, et puisque $U(x,y) > 0$ pour tous $x, y \in E$, on voit aussi qu'il y a une seule classe de récurrence. Le reste découle du théorème, à l'exception de la dernière assertion : si E est fini et si on suppose que tous les états sont transitoires, on a

$$\mathbb{P}_x \text{ p.s. },\quad \sum_{y \in E} N_y < \infty$$

ce qui est absurde puisque

$$\sum_{y\in E} N_y = \sum_{y\in E}\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=y\}} = \sum_{n=0}^{\infty}\sum_{y\in E} \mathbf{1}_{\{X_n=y\}} = \infty.$$

□

Une chaîne de Markov irréductible dont les états sont récurrents sera dite récurrente irréductible.

Exemples. Nous reprenons maintenant les différents exemples introduits ci-dessus pour discuter dans chaque cas la classification des états. Avant cela, insistons sur le fait que les résultats obtenus pour la chaîne de Markov canonique se traduisent immédiatement pour une chaîne de Markov quelconque $(Y_n)_{n\in\mathbb{N}}$ de transition Q (et inversement). Par exemple, si $Y_0 = y$, en notant $N_x^Y = \sum_{n=0}^{\infty} \mathbf{1}_{\{Y_n=x\}}$, on a pour tout $k \in \overline{\mathbb{N}}$,

$$P(N_x^Y = k) = \mathbb{P}_y(N_x = k))$$

puisque le terme de gauche s'écrit aussi bien

$$P((Y_n)_{n\in\mathbb{N}} \in B)$$

avec $B = \{\omega \in E^{\mathbb{N}} : N_x(\omega) = k\}$, et il suffit d'utiliser la remarque (b) suivant le théorème 13.3.3.

(1) **Cas de variables aléatoires indépendantes de loi** μ. Dans ce cas $Q(x,y) = \mu(y)$. On voit facilement que y est récurrent ssi $\mu(y) > 0$, et il y a une seule classe de récurrence. La chaîne est irréductible ssi $\mu(y) > 0$ pour tout $y \in E$.

(2) **Marche aléatoire sur** $\mathbb{Z}$. On a

$$Y_n = Y_0 + \sum_{i=1}^{n} \xi_i$$

où les v.a. ξ_i, à valeurs dans $\mathbb{Z}$, sont indépendantes et de loi μ (et indépendantes de Y_0). Dans ce cas, puisque $Q(x,y) = \mu(y-x)$, on voit aisément que $U(x,y)$ est fonction de $y-x$, et donc tous les états sont du même type, récurrent ou transitoire.

Théorème 13.4.6 *Supposons* $E[|\xi_1|] < \infty$ *et soit* $m = E[\xi_1]$.

(i) *Si* $m \neq 0$, *tous les états sont transitoires.*

(ii) *Si* $m = 0$, *tous les états sont récurrents. De plus, la chaîne est irréductible ssi le sous-groupe engendré par* $\{y \in \mathbb{Z} : \mu(y) > 0\}$ *est* $\mathbb{Z}$ *tout entier.*

Preuve. (i) Si $m \neq 0$, la loi forte des grands nombres montre aussitôt que $|Y_n| \longrightarrow \infty$ p.s. et donc tous les états sont transitoires.
(ii) Supposons que $m = 0$ et que 0 est transitoire, donc $U(0,0) < \infty$. Nous allons voir que ceci conduit à une contradiction. Sans perte de généralité, on suppose dans la suite que $Y_0 = 0$. On observe que, pour tout $x \in \mathbb{Z}$,

$$U(0,x) \leq U(x,x) = U(0,0)$$

la première inégalité découlant de la proposition 13.4.2(iii). En conséquence, pour tout $n \geq 1$,

$$\sum_{|x| \leq n} U(0,x) \leq (2n+1)U(0,0) \leq Cn \tag{13.2}$$

avec $C = 3U(0,0) < \infty$.

D'autre part, on sait que $n^{-1}Y_n$ converge p.s., donc aussi en probabilité, vers 0. Si on pose $\varepsilon = (4C)^{-1}$, on peut trouver N assez grand pour que, pour tout $n \geq N$,

$$P(|Y_n| \leq \varepsilon n) > \frac{1}{2},$$

ou de manière équivalente,

$$\sum_{|x| \leq \varepsilon n} Q_n(0,x) > \frac{1}{2}.$$

Si $n \geq p \geq N$, on a aussi

$$\sum_{|x| \leq \varepsilon n} Q_p(0,x) \geq \sum_{|x| \leq \varepsilon p} Q_p(0,x) > \frac{1}{2}$$

puis en sommant sur p,

$$\sum_{|x| \leq \varepsilon n} U(0,x) \geq \sum_{p=N}^{n} \sum_{|x| \leq \varepsilon p} Q_p(0,x) > \frac{n-N}{2}.$$

Mais d'autre part, d'après (13.2), si $\varepsilon n \geq 1$,

$$\sum_{|x|\leq \varepsilon n} U(0,x) \leq C\varepsilon n = \frac{n}{4}.$$

On obtient une contradiction dès que n est assez grand.

Il reste à établir la dernière assertion. Notons G le sous-groupe engendré par $\{x \in \mathbb{Z} : \mu(x) > 0\}$. Il est immédiat que

$$P(Y_n \in G, \forall n \in \mathbb{N}) = 1$$

(rappelons que nous avons pris $Y_0 = 0$). Cela montre que si $G \neq \mathbb{Z}$, la chaîne n'est pas irréductible. Inversement, supposons que $G = \mathbb{Z}$. Alors, notons

$$H = \{x \in \mathbb{Z} : U(0,x) > 0\}$$

et observons que H est un sous-groupe de $\mathbb{Z}$:

- si $x, y \in H$, l'inégalité
$$Q_{n+p}(0, x+y) \geq Q_n(0,x)\, Q_p(x, x+y) = Q_n(0,x)\, Q_p(0,y)$$
montre que $x + y \in H$;
- si $x \in H$, comme 0 est récurrent, la condition $U(0,x) > 0$ entraîne $U(x,0) > 0$ (lemme 13.4.3) et puisque $U(x,0) = U(0,-x)$ on a bien $-x \in H$.

Finalement, puisque H contient $\{x \in \mathbb{Z} : \mu(x) > 0\}$, on a forcément $H = \mathbb{Z}$. □

Par exemple, si $\mu = \frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_2$, tous les états sont récurrents, mais il y a deux classes de récurrence, les entiers pairs et les entiers impairs.

(3) **Marche aléatoire sur un graphe**. On considère ici le cas d'un graphe fini : E est fini et A est un sous-ensemble de $\mathcal{P}_2(E)$ tel que, pour tout $x \in E$, $A_x := \{y \in E : \{x, y\} \in A\}$ est non vide. Le graphe est dit connexe si pour tous $x, y \in E$, on peut trouver un entier $p \geq 0$ et des élements $x_0 = x, x_1, \ldots, x_{p-1}, x_p = y$ de E tels que $\{x_{i-1}, x_i\} \in A$ pour tout $i \in \{1, \ldots, p\}$.

Proposition 13.4.7 *La marche aléatoire simple sur un graphe fini connexe est récurrente irréductible.*

Preuve. Le caractère irréductible de la chaîne découle de la connexité du graphe. Il suffit ensuite d'appliquer le corollaire 13.4.5. □

(4) **Processus de branchement**. Dans ce cas $E = \mathbb{N}$ et $Q(x,y) = \mu^{*x}(y)$. On remarque que l'état 0 est toujours absorbant, au sens où

$$\mathbb{P}_0(\forall n \in \mathbb{N}\,,\, X_n = 0) = 1.$$

En conséquence 0 est aussi récurrent.

Dans la proposition suivante, nous écartons le cas trivial $\mu = \delta_1$, où tous les états sont absorbants.

Proposition 13.4.8 0 *est le seul état récurrent. En conséquence, on a p.s.*

- *ou bien* $\exists N : \forall n \geq N \, , \, X_n = 0$.
- *ou bien* $X_n \longrightarrow +\infty$ *quand* $n \to \infty$.

Remarque. On a vu dans le chapitre précédent que le premier cas se produit avec probabilité 1 si $m = \sum k\mu(k) \leq 1$, et que le second cas se produit avec probabilité strictement positive si $m > 1$ (sous l'hypothèse supplémentaire que μ a un moment d'ordre 2).

Preuve. Supposons d'abord que $\mu(0) > 0$. Si $x \geq 1$, $U(x,0) \geq \mathbb{P}_x(X_1 = 0) = \mu(0)^x > 0$ alors que $U(0,x) = 0$. Cela n'est possible que si x est transitoire. Traitons ensuite le cas où $\mu(0) = 0$. Comme nous excluons le cas $\mu = \delta_1$, il existe alors $k \geq 2$ tel que $\mu(k) > 0$. Alors, pour tout $x \geq 1$, $\mathbb{P}_x(X_1 > x) > 0$, ce qui entraîne qu'il existe $y > x$ tel que $U(x,y) > 0$. Comme on a clairement $U(y,x) = 0$, on conclut encore que x est transitoire. Les autres assertions découlent maintenant du théorème 13.4.4. □

13.5 Mesures invariantes

Définition 13.5.1 *Soit* μ *une mesure positive sur* E, *telle que* $\mu(x) < \infty$ *pour tout* $x \in E$ *et* μ *n'est pas la mesure identiquement nulle. On dit que* μ *est invariante pour la matrice de transition* Q *(ou simplement invariante s'il n'y a pas ambiguïté) si*

$$\forall y \in E \, , \quad \mu(y) = \sum_{x \in E} \mu(x) Q(x,y).$$

Sous forme matricielle, la condition d'invariance s'écrit $\mu Q = \mu$. Puisque pour tout n, $Q_n = (Q)^n$, on peut itérer cette relation et obtenir que $\mu Q_n = \mu$ pour tout $n \in \mathbb{N}$.

Interprétation. Supposons de plus que $\mu(E) < \infty$ (ce qui sera toujours le cas si E est fini). Quitte à remplacer μ par $\mu(E)^{-1}\mu$, on peut supposer $\mu(E) = 1$. Alors, pour toute fonction $f : E \longrightarrow \mathbb{R}_+$,

$$\mathbb{E}_\mu[f(X_1)] = \sum_{x \in E} \mu(x) \sum_{y \in E} Q(x,y) f(y) = \sum_{y \in E} f(y) \sum_{x \in E} \mu(x) Q(x,y) = \sum_{y \in E} \mu(y) f(y)$$

ce qui montre que sous $\mathbb{P}_\mu$, X_1 a même loi μ que X_0. En utilisant la relation $\mu Q_n = Q$, on obtient de même que pour tout $n \in \mathbb{N}$ la loi de X_n sous $\mathbb{P}_\mu$ est μ. Plus précisément, pour toute fonction $F : \Omega \longrightarrow \mathbb{R}_+$ mesurable,

$$\mathbb{E}_\mu[F \circ \theta_1] = \mathbb{E}_\mu[\mathbb{E}_{X_1}[F]] = \sum_{x \in E} \mu(x) \, \mathbb{E}_x[F] = \mathbb{E}_\mu[F]$$

ce qui montre que sous $\mathbb{P}_\mu$, $(X_{1+n})_{n \in \mathbb{N}}$ a même loi que $(X_n)_{n \in \mathbb{N}}$ (et de même, pour tout entier $k \geq 0$, $(X_{k+n})_{n \in \mathbb{N}}$ a même loi que $(X_n)_{n \in \mathbb{N}}$).

Exemple. Pour toute marche aléatoire sur $\mathbb{Z}^d$ ($Q(x,y) = \gamma(y-x)$ ne dépend que la différence $y - x$), on vérifie immédiatement que la mesure de comptage sur $\mathbb{Z}^d$ est invariante.

Définition 13.5.2 *Soit μ une mesure positive non triviale sur E, telle que $\mu(x) < \infty$ pour tout $x \in E$. On dit que μ est réversible si*

$$\forall x, y \in E\,, \quad \mu(x)Q(x,y) = \mu(y)Q(y,x).$$

Proposition 13.5.1 *Toute mesure réversible est invariante.*

Preuve. Si μ est réversible,

$$\sum_{x \in E} \mu(x)Q(x,y) = \sum_{x \in E} \mu(y)Q(y,x) = \mu(y).$$

□

En revanche, il existe des mesures invariantes qui ne sont pas réversibles : nous avons vu que la mesure de comptage est invariante pour toute marche aléatoire sur $\mathbb{Z}^d$, cependant elle n'est réversible que si la loi de saut γ est symétrique ($\gamma(x) = \gamma(-x)$).

Exemples. (a) **Pile ou face biaisé**. C'est la marche aléatoire sur $\mathbb{Z}$ de matrice de transition

$$\begin{array}{l} Q(i, i+1) = p \\ Q(i, i-1) = q = 1 - p \end{array}$$

où $p \in]0,1[$. Dans ce cas, on vérifie aisément que la mesure

$$\mu(i) = (\frac{p}{q})^i\,, \qquad i \in \mathbb{Z}$$

est réversible, donc invariante. Remarquons que μ est différente de la mesure de comptage (qui est aussi invariante) sauf dans le cas $p = 1/2$.

(b) **Marche aléatoire sur un graphe**. La mesure

$$\mu(x) = \mathrm{Card}(A_x)$$

est réversible. En effet, si $\{x, y\} \in A$,

$$\mu(x)Q(x,y) = \mathrm{Card}(A_x)\,\frac{1}{\mathrm{Card}(A_x)} = 1 = \mu(y)Q(y,x).$$

(c) **Modèle d'urne d'Ehrenfest**. C'est la chaîne de Markov dans $\{0, 1, \ldots, k\}$ de matrice de transition

$$\begin{array}{ll} Q(j, j+1) = \frac{k-j}{k} & \text{si } 0 \le j \le k-1 \\ Q(j, j-1) = \frac{j}{k} & \text{si } 1 \le j \le k. \end{array}$$

Une mesure μ est réversible ssi

$$\mu(j)\,\frac{k-j}{k} = \mu(j+1)\,\frac{j+1}{k}$$

pour tout $0 \le j \le k-1$. On trouve aisément que

$$\mu(j) = C_k^j$$

convient.

Théorème 13.5.2 *Soit x un point récurrent. La formule*

$$\mu(y) = \mathbb{E}_x\Big[\sum_{k=0}^{H_x-1} \mathbf{1}_{\{X_k=y\}}\Big]$$

définit une mesure invariante. De plus, $\mu(y) > 0$ ssi y appartient à la classe de récurrence de x.

Preuve. Remarquons d'abord que si y n'est pas dans la classe de récurrence de x on a $\mathbb{E}_x[N_y] = U(x,y) = 0$, et donc a fortiori $\mu(y) = 0$.

Ensuite, on écrit pour tout $y \in E$,

$$\begin{aligned}
\mu(y) &= \mathbb{E}_x\Big[\sum_{k=1}^{H_x} \mathbf{1}_{\{X_k=y\}}\Big] \\
&= \sum_{z\in E} \mathbb{E}_x\Big[\sum_{k=1}^{H_x} \mathbf{1}_{\{X_{k-1}=z,\, X_k=y\}}\Big] \\
&= \sum_{z\in E} \sum_{k=1}^{\infty} \mathbb{E}_x\Big[\mathbf{1}_{\{k\le H_x,\, X_{k-1}=z\}}\, \mathbf{1}_{\{X_k=y\}}\Big] \\
&= \sum_{z\in E} \sum_{k=1}^{\infty} \mathbb{E}_x\Big[\mathbf{1}_{\{k\le H_x,\, X_{k-1}=z\}}\Big]\, Q(z,y) \\
&= \sum_{z\in E} \mathbb{E}_x\Big[\sum_{k=1}^{H_x} \mathbf{1}_{\{X_{k-1}=z\}}\Big]\, Q(z,y) \\
&= \sum_{z\in E} \mu(z) Q(z,y).
\end{aligned}$$

Dans la quatrième égalité, on a utilisé le fait que l'événement $\{k \le H_x,\, X_{k-1} = z\}$ est $\mathcal{F}_{k-1}$-mesurable pour appliquer la propriété de Markov à l'instant $k-1$.

On a obtenu l'identité $\mu Q = \mu$, qu'on peut itérer pour avoir $\mu Q_n = \mu$ pour tout entier $n \ge 0$. En particulier, pour tout entier $n \ge 0$,

$$\mu(x) = 1 = \sum_{z\in E} \mu(z) Q_n(z,x).$$

Soit y un point de la classe de récurrence de x. Alors, il existe $n \ge 0$ tel que $Q_n(y,x) > 0$, et la formule précédente montre que $\mu(y) < \infty$. On peut aussi trouver m tel que $Q_m(x,y) > 0$, et on a

$$\mu(y) = \sum_{z\in E} \mu(z) Q_m(z,y) \ge Q_m(x,y) > 0.$$

Remarque. S'il existe plusieurs classes de récurrence R_i, $i \in I$, alors en choisissant pour chaque $i \in I$ un point $x_i \in R_i$ et en posant

$$\mu_i(y) = \mathbb{E}_{x_i}\Big[\sum_{k=0}^{H_{x_i}-1} \mathbf{1}_{\{X_k=y\}}\Big]$$

on construit des mesures invariantes à supports disjoints.

Théorème 13.5.3 *Supposons la chaîne récurrente irréductible. Alors la mesure invariante est unique à une constante multiplicative près.*

Preuve. Soit μ une mesure invariante. On montre par récurrence que, pour tout entier $p \geq 0$, pour tous $x, y \in E$,

$$\mu(y) \geq \mu(x)\,\mathbb{E}_x\Big[\sum_{k=0}^{p\wedge(H_x-1)} \mathbf{1}_{\{X_k=y\}}\Big]. \tag{13.3}$$

D'abord, si $y = x$, l'inégalité est immédiate (avec même une égalité). On suppose donc $y \neq x$. Si $p = 0$, l'inégalité (13.3) est triviale. On suppose que (13.3) est vraie à l'ordre p. Alors,

$$\begin{aligned}
\mu(y) &= \sum_{z\in E} \mu(z)\,Q(z,y) \\
&\geq \mu(x)\sum_{z\in E}\mathbb{E}_x\Big[\sum_{k=0}^{p\wedge(H_x-1)} \mathbf{1}_{\{X_k=z\}}\Big]Q(z,y) \\
&= \mu(x)\sum_{z\in E}\sum_{k=0}^{p}\mathbb{E}_x\Big[\mathbf{1}_{\{X_k=z,\,k\leq H_x-1\}}\Big]Q(z,y) \\
&= \mu(x)\sum_{z\in E}\sum_{k=0}^{p}\mathbb{E}_x\Big[\mathbf{1}_{\{X_k=z,\,k\leq H_x-1\}}\,\mathbf{1}_{\{X_{k+1}=y\}}\Big] \\
&= \mu(x)\mathbb{E}_x\Big[\sum_{k=0}^{p\wedge(H_x-1)} \mathbf{1}_{\{X_{k+1}=y\}}\Big] \\
&= \mu(x)\mathbb{E}_x\Big[\sum_{k=1}^{(p+1)\wedge H_x} \mathbf{1}_{\{X_k=y\}}\Big],
\end{aligned}$$

ce qui donne le résultat voulu à l'ordre $p+1$. De manière analogue à la preuve du théorème précédent, on a utilisé le fait que l'événement $\{X_k = z,\, k \leq H_x - 1\}$ est $\mathcal{F}_k$-mesurable pour appliquer la propriété de Markov à l'instant k.

En faisant tendre p vers $+\infty$ dans (13.3) on trouve

$$\mu(y) \geq \mu(x)\,\mathbb{E}_x\Big[\sum_{k=0}^{H_x-1} \mathbf{1}_{\{X_k=y\}}\Big].$$

Fixons $x \in E$. La mesure

$$\nu_x(y) = \mathbb{E}_x\Big[\sum_{k=0}^{H_x-1} \mathbf{1}_{\{X_k=y\}}\Big]$$

est invariante (théorème 13.5.2), et on a $\mu(y) \geq \mu(x)\nu_x(y)$ pour tout $y \in E$. Donc, pour tout $n \geq 1$,

$$\mu(x) = \sum_{z\in E} \mu(z)Q_n(z,x) \geq \sum_{z\in E} \mu(x)\nu_x(z)Q_n(z,x) = \mu(x)\nu_x(x) = \mu(x),$$

ce qui montre que l'égalité $\mu(z) = \mu(x)\nu_x(z)$ a lieu pour tout z tel que $Q_n(z,x) > 0$. L'irréductibilité assure que pour tout $z \in E$ on peut trouver un entier n tel que $Q_n(z,x) > 0$, et on conlut donc que $\mu = \mu(x)\nu_x$, ce qui termine la preuve. □

Corollaire 13.5.4 *Supposons la chaîne récurrente irréductible. Alors :*

(i) *Ou bien il existe une mesure de probabilité invariante μ, et on a pour tout $x \in E$,*

$$\mathbb{E}_x[H_x] = \frac{1}{\mu(x)}.$$

(ii) *Ou bien toute mesure invariante a une masse totale infinie, et on a pour tout $x \in E$,*

$$\mathbb{E}_x[H_x] = \infty.$$

La chaîne est dite récurrente positive dans le cas (i) *et récurrente nulle dans le cas* (ii).

Remarque. Si E est fini seul le cas (i) se produit.

Preuve. D'après le théorème 13.5.3, toutes les mesures invariantes sont proportionnelles. Donc ou bien elles sont toutes de masse totale infinie (cas (ii)) ou bien elles sont toutes finies, et on peut normaliser pour en trouver une qui soit une mesure de probabilité (cas (i)). Dans le cas (i), soit μ l'unique mesure de probabilité invariante et soit $x \in E$. Alors, si ν_x désigne la mesure invariante fournie par le théorème 13.5.2,

$$\nu_x(y) = \mathbb{E}_x\Big[\sum_{k=0}^{H_x-1} \mathbf{1}_{\{X_k=y\}}\Big],$$

μ est proportionnelle à ν_x : $\mu = C\nu_x$ avec $C > 0$. En écrivant $1 = \mu(E) = C\,\nu_x(E)$, on trouve $C = (\nu_x(E))^{-1}$, d'où

$$\mu(x) = \frac{\nu_x(x)}{\nu_x(E)} = \frac{1}{\nu_x(E)}.$$

Or

$$\nu_x(E) = \sum_{y\in E} \mathbb{E}_x\Big[\sum_{k=0}^{H_x-1} \mathbf{1}_{\{X_k=y\}}\Big] = \mathbb{E}_x\Big[\sum_{k=0}^{H_x-1}\Big(\sum_{y\in E} \mathbf{1}_{\{X_k=y\}}\Big)\Big] = \mathbb{E}_x[H_x].$$

Dans le cas (ii), ν_x est infinie, et donc, par le même calcul,

$$\mathbb{E}_x[H_x] = \nu_x(E) = \infty.$$

Proposition 13.5.5 *Supposons la chaîne irréductible. S'il existe une mesure invariante finie, la chaîne est récurrente (et donc récurrente positive).*

Preuve. Soit γ une mesure invariante finie, et soit $y \in E$ tel que $\gamma(y) > 0$. Pour tout $x \in E$, la proposition 13.4.2(iii) donne l'inégalité

$$\sum_{n=0}^{\infty} Q_n(x,y) = U(x,y) \leq U(y,y).$$

On multiplie les deux membres de cette inégalité par $\gamma(x)$ et on somme sur toutes les valeurs de $x \in E$. Il vient

$$\sum_{n=0}^{\infty} \gamma Q_n(y) \leq \gamma(E)\, U(y,y).$$

Puisque γ est invariante on a $\gamma Q_n(y) = \gamma(y) > 0$ pour tout $n \geq 0$. On conclut donc que

$$\gamma(E)\, U(y,y) = \infty.$$

Comme $\gamma(E) < \infty$, cela entraîne que $U(y,y) = \infty$. Donc y est récurrent et puisque la chaîne est irréductible elle est récurrente (corollaire 13.4.5). □

Remarque. L'existence d'une mesure invariante infinie ne permet pas de conclure : considérer par exemple le pile ou face biaisé (exemple (1) ci-dessus après la proposition 13.5.1) qui n'est récurrent que si $p = 1/2$.

Exemple. Soit $p \in]0,1[$. Considérons la chaîne de Markov sur $E = \mathbb{N}$ de matrice de transition

$$\begin{aligned} &Q(k,k+1) = p\ ,\ Q(k,k-1) = 1-p\ , \qquad \text{si } k \geq 1, \\ &Q(0,1) = 1. \end{aligned}$$

Cette chaîne est irréductible. De plus on vérifie immédiatement que la mesure μ définie par

$$\begin{aligned} &\mu(k) = \Big(\frac{p}{1-p}\Big)^{k-1}, \qquad \text{si } k \geq 1, \\ &\mu(0) = 1-p\ , \end{aligned}$$

est réversible donc invariante.

Si $p < \frac{1}{2}$, la mesure μ est finie, et la proposition 13.5.5 entraîne que la chaîne est récurrente positive. (Exercice : Montrer que la chaîne est récurrente nulle si $p = \frac{1}{2}$, et transitoire si $p > \frac{1}{2}$.)

13.6 Comportement asymptotique

Nous continuons à considérer la chaîne de Markov canonique associée à une matrice de transition Q.

Théorème 13.6.1 *Supposons la chaîne récurrente irréductible, et soit μ une mesure invariante. Soient f et g deux fonctions positives sur E telles que $\int f\,d\mu < \infty$ et $0 < \int g\,d\mu < \infty$. Alors, pour tout $x \in E$ on a $\mathbb{P}_x$ p.s.*

$$\frac{\sum_{k=0}^{n} f(X_k)}{\sum_{k=0}^{n} g(X_k)} \underset{n\to\infty}{\longrightarrow} \frac{\int f\,d\mu}{\int g\,d\mu}.$$

Remarque. Le résultat reste vrai si $\mu(f) = \infty$. Il suffit d'utiliser un argument de comparaison en écrivant $f = \lim \uparrow f_k$, avec des fonctions positives f_k telles que $\int f_k d\mu < \infty$.

Corollaire 13.6.2 *Si la chaîne de Markov est irréductible et récurrente positive, et si μ désigne l'unique probabilité invariante, on a $\mathbb{P}_x$ p.s.*

$$\frac{1}{n}\sum_{k=0}^{n} f(X_k) \underset{n\to\infty}{\longrightarrow} \int f\,d\mu.$$

Le corollaire découle immédiatement du théorème en prenant $g = 1$ dans l'énoncé.

Preuve du théorème 13.6.1. On définit les temps d'arrêt

$$T_0 = 0\ ,\ T_1 = H_x$$

et par récurrence

$$T_{n+1} = \inf\{k > T_n : X_k = x\}.$$

Le temps T_n est l'instant du n-ième retour en x de la chaîne. Puisque l'état x est récurrent, tous ces temps d'arrêt sont finis p.s. On pose aussi pour tout $k \geq 0$,

$$Z_k(f) = \sum_{n=T_k}^{T_{k+1}-1} f(X_n).$$

Lemme 13.6.3 *Les v.a. $Z_k(f)$, $k = 0, 1, 2, \ldots$, sont indépendantes et de même loi.*

Preuve. Soient $g_0, g_1, g_2, \ldots$ des fonctions mesurables bornées sur $\mathbb{R}_+$. Il suffit de montrer que, pour tout entier $k \geq 0$, on a

$$\mathbb{E}_x\Big[\prod_{i=0}^{k} g_i(Z_i(f))\Big] = \prod_{i=0}^{k} \mathbb{E}_x[g_i(Z_0(f))].$$

On démontre cette identité par récurrence sur k. Pour $k = 0$ il n'y a rien à montrer. Pour passer de l'ordre $k-1$ à l'ordre k, on observe que :

- les v.a. $Z_0(f), Z_1(f), \ldots, Z_{k-1}(f)$ sont $\mathcal{F}_{T_k}$-mesurables (exercice !);
- la suite translatée $\theta_{T_k}(\omega)$ est indépendante de $\mathcal{F}_{T_k}$ et de loi $\mathbb{P}_x$, d'après le corollaire 13.3.6;
- on a $Z_k(f) = Z_0(f) \circ \theta_{T_k}$, par construction.

Il découle de tout ceci que

$$\mathbb{E}_x\Big[\prod_{i=0}^{k} g_i(Z_i(f))\Big] = \mathbb{E}_x\Big[\Big(\prod_{i=0}^{k-1} g_i(Z_i(f))\Big) g_k(Z_0(f)\circ\theta_{T_k})\Big] = \mathbb{E}_x\Big[\prod_{i=0}^{k-1} g_i(Z_i(f))\Big]\,\mathbb{E}_x[g_k(Z_0(f))],$$

d'où le résultat voulu à l'ordre k. □

Nous revenons à la preuve du théorème. Si ν_x désigne comme précédemment la mesure invariante construite dans le théorème 13.5.2, on a $\mu = \mu(x)\nu_x$ puisque $\nu_x(x) = 1$ et que toutes les mesures invariantes sont proportionnelles (théorème 13.5.3). On observe alors que

$$\mathbb{E}_x[Z_0(f)] = \mathbb{E}_x\Big[\sum_{k=0}^{H_x-1}\sum_{y\in E} f(y)\,\mathbf{1}_{\{X_k=y\}}\Big] = \sum_{y\in E} f(y)\,\nu_x(y) = \frac{\int f\,d\mu}{\mu(x)}.$$

Le lemme 13.6.3 et la loi forte des grands nombres montrent ensuite que $\mathbb{P}_x$ p.s.

$$\frac{1}{n}\sum_{k=0}^{n-1} Z_k(f) \underset{n\to\infty}{\longrightarrow} \frac{\int f\,d\mu}{\mu(x)}. \tag{13.4}$$

Pour tout entier n, notons $N_x(n)$ le nombre de retours en x effectués par la chaîne avant l'instant n, de sorte que $T_{N_x(n)} \le n < T_{N_x(n)+1}$. En écrivant

$$\frac{\sum_{k=0}^{T_{N_x(n)}-1} f(X_k)}{N_x(n)} \le \frac{\sum_{k=0}^{n} f(X_k)}{N_x(n)} \le \frac{\sum_{k=0}^{T_{N_x(n)+1}-1} f(X_k)}{N_x(n)}$$

ce qui équivaut à

$$\frac{\sum_{j=0}^{N_x(n)-1} Z_j(f)}{N_x(n)} \le \frac{\sum_{k=0}^{n} f(X_k)}{N_x(n)} \le \frac{\sum_{j=0}^{N_x(n)} Z_j(f)}{N_x(n)}$$

on déduit de la convergence (13.4) que $\mathbb{P}_x$ p.s.

$$\frac{1}{N_x(n)}\sum_{k=0}^{n} f(X_k) \underset{n\to\infty}{\longrightarrow} \frac{\int f\,d\mu}{\mu(x)}.$$

Il suffit ensuite d'utiliser le même résultat avec f remplacée par g pour finir la preuve. □

Corollaire 13.6.4 *Supposons la chaîne récurrente irréductible. Alors, pour tout $x \in E$,*

(i) *dans le cas récurrent positif,*

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_{\{X_k=x\}} \xrightarrow[n\to\infty]{\text{p.s.}} \mu(x),$$

où μ est l'unique probabilité invariante;

(ii) *dans le cas récurrent nul,*

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_{\{X_k=x\}} \xrightarrow[n\to\infty]{\text{p.s.}} 0.$$

Dans les deux cas la convergence a lieu pour toute loi initiale de la chaîne.

Définition 13.6.1 *Soit x un point récurrent, et*

$$L_x = \{n \geq 0 : Q_n(x,x) > 0\}.$$

La période de x, notée $d(x)$, est le PGCD de L_x.

Remarque. Puisque L_x est stable par addition ($Q_{n+m}(x,x) \geq Q_n(x,x)Q_m(x,x)$), le sous groupe engendré par L_x est $L_x - L_x = d(x)\mathbb{Z}$.

Proposition 13.6.5 *Supposons la chaîne récurrente irréductible.*

(i) *Tous les points ont la même période, appelée la période de la chaîne et notée d.*

(ii) *Si $d = 1$ (la chaîne est alors dite apériodique), pour tous $x, y \in E$, il existe un entier n_0 tel que $Q_n(x,y) > 0$ pour tout $n \geq n_0$.*

Preuve. (i) Soient $x, y \in E$. Puisque la chaîne est irréductible, il existe deux entiers n_1 et n_2 tels que $Q_{n_1}(x,y) > 0$ et $Q_{n_2}(y,x) > 0$. Mais alors, si $n \in L_x$, on a $n_1 + n + n_2 \in L_y$, ce qui entraîne que $L_x - L_x \subset L_y - L_y$ et donc $d(y)$ divise $d(x)$. Par symétrie on a $d(y) = d(x)$.
(ii) Clairement, il suffit de traiter le cas où $y = x$. Puisque $d(x) = 1$, on peut trouver deux entiers $n_1, m_1 \geq 0$ tels que $1 = n_1 - m_1$ et

$$Q_{n_1}(x,x) > 0,\ Q_{m_1}(x,x) > 0.$$

Si $m_1 = 0$, donc $n_1 = 1$ le résultat est évident avec $n_0 = 0$. Si $m_1 \geq 1$, alors, pour tout $j \in \{0, 1, \ldots, m_1 - 1\}$, on a

$$Q_{m_1^2+j}(x,x) = Q_{jn_1+(m_1-j)m_1}(x,x) > 0.$$

Il en découle que, si $n_0 = m_1^2$ on a pour tout entier $j \geq 0$,

$$Q_{n_0+j}(x,x) > 0.$$

Théorème 13.6.6 *Supposons la chaîne irréductible, récurrente positive et apériodique. Alors, si μ désigne l'unique probabilité invariante, on a pour tout $x \in E$,*

$$\sum_{y \in E} |\mathbb{P}_x(X_n = y) - \mu(y)| \underset{n\to\infty}{\longrightarrow} 0.$$

Preuve. La formule

$$\overline{Q}((x_1,x_2),(y_1,y_2)) = Q(x_1,y_1)Q(x_2,y_2)$$

définit une matrice stochastique sur le $E \times E$. On note $((X_n^1, X_n^2)_{n\in\mathbb{N}}, (\overline{P}_{(x_1,x_2)})_{(x_1,x_2)\in E\times E})$ la chaîne de Markov canonique associée.

Remarquons que Q est irréductible : si $(x_1,x_2),(y_1,y_2) \in E \times E$, la proposition 13.6.5(ii) permet de trouver deux entiers n_1 et n_2 tels que $Q_n(x_1,y_1) > 0$ pour tout $n \geq n_1$, et $Q_n(x_2,y_2) > 0$ pour tout $n \geq n_2$. Si $n \geq n_1 \vee n_2$, on a par définition $\overline{Q}_n((x_1,x_2),(y_1,y_2)) > 0$.

De plus la mesure produit $\mu \otimes \mu$ est invariante pour $\overline{Q}$:

$$\begin{aligned}\sum_{(x_1,x_2)\in E\times E} \mu(x_1)\mu(x_2)Q(x_1,y_1)Q(x_2,y_2) &= \sum_{x_1\in E}\mu(x_1)Q(x_1,y_1)\sum_{x_2\in E}\mu(x_2)Q(x_2,y_2)\\ &= \mu(y_1)\mu(y_2).\end{aligned}$$

La proposition 13.5.5 permet de conclure que la chaîne (X_n^1, X_n^2) est récurrente positive. Observons maintenant que

$$\mathbb{P}_x(X_n = y) - \mu(y) = \overline{\mathbb{P}}_{\mu\otimes\delta_x}(X_n^2 = y) - \overline{\mathbb{P}}_{\mu\otimes\delta_x}(X_n^1 = y) = \overline{\mathbb{E}}_{\mu\otimes\delta_x}[\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}}].$$

Introduisons le temps d'arrêt $T = \inf\{n \geq 0 : X_n^1 = X_n^2\}$. Alors, l'égalité précédente montre que

$$\begin{aligned}\mathbb{P}_x(X_n = y) - \mu(y) &= \overline{\mathbb{E}}_{\mu\otimes\delta_x}[\mathbf{1}_{\{T>n\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})]\\ &\quad+\sum_{k=0}^{n}\sum_{z\in E}\overline{\mathbb{E}}_{\mu\otimes\delta_x}[\mathbf{1}_{\{T=k,X_k^1=X_k^2=z\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})]. \quad (13.5)\end{aligned}$$

Mais, pour tout $k \in \{0, 1, \dots, n\}$ et tout $z \in E$, la propriété de Markov entraîne que

$$\begin{aligned}\overline{\mathbb{E}}_{\mu\otimes\delta_x}[\mathbf{1}_{\{T=k,X_k^1=X_k^2=z\}}\mathbf{1}_{\{X_n^2=y\}}] &= \overline{\mathbb{E}}_{\mu\otimes\delta_x}[\mathbf{1}_{\{T=k,X_k^1=X_k^2=z\}}]\,Q_{n-k}(z,y)\\ &= \overline{\mathbb{E}}_{\mu\otimes\delta_x}[\mathbf{1}_{\{T=k,X_k^1=X_k^2=z\}}\mathbf{1}_{\{X_n^1=y\}}],\end{aligned}$$

et donc le deuxième terme de la somme dans (13.5) est nul. On obtient ainsi que

$$\begin{aligned}\sum_{y\in E}|\mathbb{P}_x(X_n = y) - \mu(y)| &= \sum_{y\in E}|\overline{\mathbb{E}}_{\mu\otimes\delta_x}[\mathbf{1}_{\{T>n\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})]|\\ &\leq \sum_{y\in E}\overline{\mathbb{E}}_{\mu\otimes\delta_x}[\mathbf{1}_{\{T>n\}}(\mathbf{1}_{\{X_n^2=y\}} + \mathbf{1}_{\{X_n^1=y\}})]\\ &= 2\,\overline{\mathbb{P}}_{\mu\otimes\delta_x}(T > n),\end{aligned}$$

qui tend vers 0 quand $n \to \infty$, grâce à la récurrence de la chaîne (X_n^1, X_n^2). □

13.7 Martingales et chaînes de Markov

On considère toujours la chaîne de Markov canonique de matrice de transition Q.

Définition 13.7.1 *Une fonction $f : E \longrightarrow \mathbb{R}_+$ est dite harmonique (resp. surharmonique) si on a pour tout $x \in E$,*

$$f(x) = Qf(x) \qquad (\textit{resp. } f(x) \geq Qf(x)).$$

Plus généralement, si $F \subset E$, on dit que f est harmonique sur F (resp. surharmonique sur F) si la propriété $f(x) = Qf(x)$ (resp. $f(x) \geq Qf(x)$) est vraie pour $x \in F$.

Remarque. On pourrait considérer plus généralement des fonctions harmoniques ou surharmoniques de signe quelconque.

Proposition 13.7.1 (i) *La fonction f est harmonique (resp. surharmonique) ssi, pour tout $x \in E$, le processus $(f(X_n))_{n\in\mathbb{N}}$ est une martingale (resp. une surmartingale) sous $\mathbb{P}_x$, relativement à la filtration $(\mathcal{F}_n)$.*
(ii) *Soit $F \subset E$ et $G = E\backslash F$. On note T_G le temps d'arrêt*

$$T_G = \inf\{n \geq 0 : X_n \in G\}.$$

Alors si f est harmonique (resp. surharmonique) sur F, le processus $(f(X_{n\wedge T_G}))_{n\in\mathbb{N}}$ est une martingale (resp. une surmartingale) sous $\mathbb{P}_x$, pour tout $x \in F$.

Preuve. (i) Supposons d'abord f harmonique. Alors, d'après la proposition 13.1.2(i),

$$\mathbb{E}_x[f(X_{n+1}) \mid \mathcal{F}_n] = Qf(X_n) = f(X_n)$$

et en conséquence $\mathbb{E}_x[f(X_n)] = \mathbb{E}_x[f(X_0)] = f(x)$, donc $f(X_n) \in L^1$.

Inversement, supposons que $f(X_n)$ est une martingale sour $\mathbb{P}_x$. Il vient immédiatement que

$$f(x) = \mathbb{E}_x[f(X_0)] = \mathbb{E}_x[f(X_1)] = Qf(x).$$

Le cas d'une fonction surharmonique est traité de la même façon.

(ii) Traitons le cas d'une fonction harmonique. On écrit pour $x \in F$

$$\begin{aligned}
\mathbb{E}_x[f(X_{(n+1)\wedge T_G}) \mid \mathcal{F}_n] &= \mathbb{E}_x[f(X_{n+1})\,\mathbf{1}_{\{T_G>n\}} \mid \mathcal{F}_n] + \mathbb{E}_x[f(X_{T_G})\,\mathbf{1}_{\{T_G\leq n\}} \mid \mathcal{F}_n] \\
&= \mathbf{1}_{\{T_G>n\}}\,\mathbb{E}_x[f(X_{n+1}) \mid \mathcal{F}_n] + f(X_{T_G})\,\mathbf{1}_{\{T_G\leq n\}} \\
&= \mathbf{1}_{\{T_G>n\}}\,Qf(X_n) + f(X_{T_G})\,\mathbf{1}_{\{T_G\leq n\}} \\
&= \mathbf{1}_{\{T_G>n\}}\,f(X_n) + f(X_{T_G})\,\mathbf{1}_{\{T_G\leq n\}} \\
&= f(X_{n\wedge T_G})
\end{aligned}$$

On a utilisé le fait que $f(X_{T_G})\,\mathbf{1}_{\{T_G\leq n\}} = f(X_{T_G\wedge n})\,\mathbf{1}_{\{T_G\leq n\}}$ est $\mathcal{F}_n$-mesurable. □

Théorème 13.7.2 *Soit F un sous-ensemble non vide de E et $G = E\backslash F$. Soit $g : G \longrightarrow \mathbb{R}_+$ une fonction bornée.*

(i) *La fonction*

$$h(x) = \mathbb{E}_x[g(X_{T_G})\,\mathbf{1}_{\{T_G<\infty\}}], \qquad x \in E$$

est harmonique sur F.

(ii) *Supposons $T_G < \infty$, $\mathbb{P}_x$ p.s. pour tout $x \in F$. Alors la fonction h est l'unique fonction bornée sur E qui*

- *est harmonique sur F,*
- *coïncide avec g sur G.*

Preuve. (i) On remarque que si $x \in F$ on a $\mathbb{P}_x$ p.s.

$$g(X_{T_G})\,\mathbf{1}_{\{T_G<\infty\}} = g(X_{T_G}\circ\theta_1)\,\mathbf{1}_{\{T_G\circ\theta_1<\infty\}}.$$

Autrement dit, si $U(\omega) = g(X_{T_G}(\omega))\,\mathbf{1}_{\{T_G(\omega)<\infty\}}$, on a $U = U\circ\theta_1$, $\mathbb{P}_x$ p.s. Donc, pour $x \in F$, d'après le théorème 13.3.4,

$$h(x) = \mathbb{E}_x[U] = \mathbb{E}_x[U\circ\theta_1] = \mathbb{E}_x[\mathbb{E}_{X_1}[U]] = \mathbb{E}_x[h(X_1)] = Qh(x),$$

ce qui montre que h est harmonique sur F.

(ii) Il est trivial que $h(x) = g(x)$ si $x \in G$. Soit h' une autre fonction harmonique sur F, bornée sur E et coïncidant avec g sur G. Si $x \in F$, d'après la proposition 13.7.1, $Y_n = h'(X_{n\wedge T_G})$ est une martingale sous $\mathbb{P}_x$. Cette martingale est bornée, donc uniformément intégrable, et converge $\mathbb{P}_x$ p.s. vers $h'(X_{T_G}) = g(X_{T_G})$. D'après les résultats du chapitre 12, on a donc

$$h'(x) = \mathbb{E}_x[Y_0] = \mathbb{E}_x[Y_\infty] = \mathbb{E}_x[g(X_{T_G})] = h(x).$$

Exemple. *Problème de Dirichlet discret.* Soit F une partie finie de $\mathbb{Z}^d$. La frontière de F est

$$\partial F = \{y \in \mathbb{Z}^d\backslash F : \exists x \in F,\ |y-x| = 1\}.$$

On note $\overline{F} = F \cup \partial F$.

Une fonction h définie sur $\overline{F}$ est dite harmonique (au sens discret) sur F si pour tout $x \in F$, $h(x)$ est égal à la moyenne des valeurs de h sur les $2d$ plus proches voisins de x. On retrouve la notion précédente en prenant comme chaîne de Markov la marche aléatoire simple sur $\mathbb{Z}^d$: $Q(x, x\pm e_j) = \frac{1}{2d}$ pour $j = 1,\ldots,d$, où $(e_1,\ldots,e_d)$ est la base canonique.

Alors, le théorème précédent conduit au résultat suivant : pour toute fonction (positive) g définie sur ∂F, la seule fonction $h : \overline{F} \longrightarrow \mathbb{R}_+$ telle que :

- h est harmonique sur F,
- $h(y) = g(y)$, $\forall y \in \partial F$,

est donnée par

$$h(x) = \mathbb{E}_x[g(X_{T_{\partial F}})]\ , \qquad x \in F,$$

où

$$T_{\partial F} = \inf\{n \geq 0 : X_n \in \partial F\}.$$

Noter que pour appliquer le théorème 13.7.2, on a a priori besoin de définir g sur $\mathbb{Z}^d\backslash F$ et non pas seulement sur ∂F : cependant le choix des valeurs de g sur $\mathbb{Z}^d\backslash\overline{F}$ n'influe pas sur les valeurs de h sur F.

Chapitre 14

Introduction au mouvement brownien

14.1 Le mouvement brownien comme limite de marches aléatoires

L'explication physique du mouvement brownien justifie le mouvement très désordonné et imprévisible d'une particule brownienne par les nombreux chocs que cette particule reçoit du milieu environnant, qui provoquent des changements de direction continuels. D'un point de vue mathématique, cela suggère de considérer le déplacement à temps discret, sur le réseau $\mathbb{Z}^d$, d'une particule ponctuelle qui à chaque instant choisit de manière indépendante du passé une nouvelle direction.

Précisément on considère une marche aléatoire $(S_n)_{n\in\mathbb{N}}$ sur $\mathbb{Z}^d$, issue de 0:

$$S_n = Y_1 + \cdots + Y_n$$

où les v.a. $Y_1, Y_2, \ldots$ sont indépendantes à valeurs dans $\mathbb{Z}^d$, et de même loi μ. On suppose que μ vérifie les propriétés suivantes :

- $\sum_{k\in\mathbb{Z}^d} |k|^2 \mu(k) < \infty$;
- $\sum_{k\in\mathbb{Z}^d} k\mu(k) = 0$ (μ est centrée).

On ajoute aussi à ces deux hypothèses principales la condition d'isotropie suivante :

- il existe une constante $\sigma > 0$ telle que pour tous $i, j \in \{1, \ldots, d\}$,

$$\sum_{k\in\mathbb{Z}^d} k_i k_j \, \mu(k) = \sigma^2 \, \delta_{ij}.$$

La marche aléatoire simple sur $\mathbb{Z}^d$ (cf chapitre précédent) vérifie ces hypothèses, avec $\sigma^2 = 1/d$, et il existe beaucoup d'autres exemples.

On va s'intéresser au comportement "global" de la fonction $k \longrightarrow S_k$ sur un "long" intervalle de temps. Pour cela on introduit le changement d'échelle suivant. Pour tout entier $n \geq 1$, pour tout réel $t \geq 0$, on pose

$$S_t^{(n)} = \frac{1}{\sqrt{n}} S_{[nt]}$$

où $[x]$ désigne la partie entière du nombre réel x.

Proposition 14.1.1 *Pour tout choix de l'entier $p \geq 1$ et des nombres réels $0 = t_0 < t_1 < \cdots < t_p$, on a*

$$(S_{t_1}^{(n)}, S_{t_2}^{(n)}, \ldots, S_{t_p}^{(n)}) \underset{n\to\infty}{\overset{\text{(loi)}}{\longrightarrow}} (U_1, U_2, \ldots, U_p)$$

et la loi limite est caractérisée comme suit:

- *les v.a. $U_1, U_2 - U_1, \ldots, U_p - U_{p-1}$ sont indépendantes;*
- *pour tout $j \in \{1, \ldots, p\}$, $U_j - U_{j-1}$ est un vecteur gaussien centré de matrice de covariance $\sigma^2(t_j - t_{j-1})\mathrm{Id}$ (par convention, $U_0 = 0$).*

Remarque. La densité de la loi limite est facile à écrire explicitement. La densité de $U_j - U_{j-1}$ est $p_{\sigma^2(t_j - t_{j-1})}(x)$, où, pour tout $a > 0$,

$$p_a(x) = \frac{1}{(2\pi a)^{d/2}} \exp\Big(-\frac{|x|^2}{2at}\Big), \quad x \in \mathbb{R}^d$$

est la densité du vecteur gaussien de covariance $a\,\mathrm{Id}$ (rappelons que les coordonnées d'un tel vecteur sont des v.a. réelles $\mathcal{N}(0,a)$ indépendantes, voir la Proposition 11.4.2 et la remarque suivant cette proposition). Grâce à l'indépendance des v.a. $U_1, U_2 - U_1, \ldots, U_p - U_{p-1}$, on obtient que la densité de $(U_1, U_2 - U_1, \ldots, U_p - U_{p-1})$ est

$$g(x_1, \ldots, x_p) = p_{\sigma^2 t_1}(x_1) p_{\sigma^2(t_2 - t_1)}(x_2) \cdots p_{\sigma^2(t_p - t_{p-1})}(x_p),$$

et par un changement de variables facile, la densité de $(U_1, U_2, \ldots, U_p)$ est

$$f(y_1, \ldots, y_p) = g(y_1, y_2 - y_1, \ldots, y_p - y_{p-1}) = p_{\sigma^2 t_1}(y_1) p_{\sigma^2(t_2 - t_1)}(y_2 - y_1) \cdots p_{\sigma^2(t_p - t_{p-1})}(y_p - y_{p-1}).$$

Preuve. Il suffit de montrer que, pour tous $\xi_1, \ldots, \xi_p \in \mathbb{R}^d$,

$$E\Big[\exp\Big(i \sum_{j=1}^p \xi_j \cdot S_{t_j}^{(n)}\Big)\Big] \underset{n\to\infty}{\longrightarrow} E\Big[\exp\Big(i \sum_{j=1}^p \xi_j \cdot U_j\Big)\Big].$$

Cela équivaut à dire que, pour tous $\eta_1, \ldots, \eta_p \in \mathbb{R}^d$,

$$E\Big[\exp\Big(i \sum_{j=1}^p \eta_j \cdot (S_{t_j}^{(n)} - S_{t_{j-1}}^{(n)})\Big)\Big] \underset{n\to\infty}{\longrightarrow} E\Big[\exp\Big(i \sum_{j=1}^p \eta_j \cdot (U_j - U_{j-1})\Big)\Big]. \tag{14.1}$$

Or on sait déjà, grâce à l'indépendance des v.a. $U_1, U_2 - U_1, \ldots, U_p - U_{p-1}$, que

$$E\Big[\exp\Big(i\sum_{j=1}^{p}\eta_j\cdot(U_j-U_{j-1})\Big)\Big] = \prod_{i=1}^{p}E\Big[\exp\Big(i\eta_j\cdot(U_j-U_{j-1})\Big)\Big] = \exp\Big(-\sum_{j=1}^{p}\frac{\sigma^2|\eta_j|^2(t_j-t_{j-1})}{2}\Big)$$

(on utilise la formule pour la transformée de Fourier de la loi gaussienne). D'autre part,

$$S^{(n)}_{t_j} - S^{(n)}_{t_{j-1}} = \frac{1}{\sqrt{n}}\sum_{k=[nt_{j-1}]+1}^{[nt_j]} Y_k$$

ce qui montre d'une part que les v.a. $S^{(n)}_{t_j} - S^{(n)}_{t_{j-1}}$, $1 \le j \le p$ sont indépendantes, d'autre part que pour chaque j fixé

$$S^{(n)}_{t_j} - S^{(n)}_{t_{j-1}} \overset{\text{(loi)}}{=} \frac{1}{\sqrt{n}}S_{[nt_j]-[nt_{j-1}]} = \frac{\sqrt{[nt_j]-[nt_{j-1}]}}{\sqrt{n}}\,\frac{1}{\sqrt{[nt_j]-[nt_{j-1}]}}\,S_{[nt_j]-[nt_{j-1}]}.$$

Grâce au théorème central limite vectoriel, cette dernière variable converge en loi quand $n \to \infty$ vers $\sqrt{t_j - t_{j-1}}\,N$, où N est un vecteur gaussien de covariance $\sigma^2\mathrm{Id}$ (on utilise aussi la propriété simple suivante : si X_n converge en loi vers X et si (a_n) est une suite de réels convergeant vers a, alors a_nX_n converge en loi vers aX). En conséquence, pour chaque j fixé,

$$E\Big[\exp\Big(i\,\eta_j\cdot(S^{(n)}_{t_j} - S^{(n)}_{t_{j-1}})\Big)\Big] \underset{n\to\infty}{\longrightarrow} E[\exp(i\sqrt{t_j-t_{j-1}}\,\eta_j\cdot N)] = \exp\Big(-\frac{\sigma^2|\eta_j|^2(t_j-t_{j-1})}{2}\Big).$$

L'indépendance des v.a. $S^{(n)}_{t_j} - S^{(n)}_{t_{j-1}}$, $1 \le j \le p$, permet maintenant de conclure au résultat recherché (14.1). □

Définition 14.1.1 *On appelle mouvement brownien (en dimension d, issu de 0) une famille $(B_t)_{t\in\mathbb{R}_+}$ de v.a. à valeurs dans $\mathbb{R}^d$, définies sur un espace de probabilité $(\Omega, \mathcal{F}, P)$, telles que :*

(P1) *On a $B_0 = 0$ p.s. De plus, pour tout choix de l'entier $p \ge 1$ et des nombres réels $0 = t_0 < t_1 < \cdots < t_p$, les v.a. $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_p} - B_{t_{p-1}}$ sont indépendantes, et, pour tout $j \in \{1, \ldots, p\}$, $B_{t_j} - B_{t_{j-1}}$ est un vecteur gaussien centré de covariance $(t_j - t_{j-1})\mathrm{Id}$.*

(P2) *Pour tout $\omega \in \Omega$, la fonction $t \to B_t(\omega)$ est continue.*

Remarques. (i) En admettant l'existence du mouvement brownien (établie ci-dessous), on peut reformuler la Proposition 14.1.1 en disant que, pour tout choix de $t_1 < \cdots < t_p$,

$$(S^{(n)}_{t_1}, S^{(n)}_{t_2}, \ldots, S^{(n)}_{t_p}) \underset{n\to\infty}{\overset{\text{(loi)}}{\longrightarrow}} (\sigma B_{t_1}, \sigma B_{t_2}, \ldots, \sigma B_{t_p}).$$

A la multiplication par le scalaire σ près, le mouvement brownien apparaît donc comme la limite continue de marches aléatoires discrètes convenablement changées d'échelle. D'une

certaine manière, cette limite correspond, pour le phénomène physique appelé mouvement brownien, au passage de l'explication microscopique aux observations macroscopiques.
(ii) Comme on l'a vu ci-dessus, la loi de $(B_{t_1}, B_{t_2}, \ldots, B_{t_p})$ est donnée par

$$P\Big((B_{t_1}, B_{t_2}, \ldots, B_{t_p}) \in A\Big) = \int_A dy_1 \ldots dy_p \, p_{t_1}(y_1) p_{t_2-t_1}(y_2 - y_1) \cdots p_{t_p-t_{p-1}}(y_p - y_{p-1}), \tag{14.2}$$

pour toute partie brélienne A de $(\mathbb{R}^d)^p$.

14.2 La construction du mouvement brownien

Théorème 14.2.1 *Le mouvement brownien existe. Autrement dit on peut construire sur un espace de probabilité convenable une famille $(B_t)_{t\in\mathbb{R}_+}$ de v.a. satisfaisant* (P1) *et* (P2)*.*

Preuve. On traite d'abord le cas $d = 1$, et dans un premier temps on va construire la famille $(B_t)_{t\in[0,1]}$. Le choix de l'espace de probabilité $(\Omega, \mathcal{F}, P)$ ne pose pas de problème : il suffit de disposer sur cet espace d'une suite de v.a. gaussiennes $\mathcal{N}(0,1)$ indépendantes (on a vu dans le chapitre précédent qu'en prenant $\Omega = [0,1]$ on pouvait construire une suite de v.a. indépendantes de loi uniforme, qu'il est facile de transformer en une suite de v.a. gaussiennes $\mathcal{N}(0,1)$ indépendantes).

Introduisons les fonctions de Haar. On pose

$$h_0(t) = 1, \quad \forall t \in [0,1]$$

puis, pour tout entier $n \geq 0$ et pour tout $k \in \{0, 1, \ldots, 2^n - 1\}$,

$$h_n^k(t) = 2^{n/2} \, \mathbf{1}_{[(2k)2^{-n-1},(2k+1)2^{-n-1}[} - 2^{n/2} \, \mathbf{1}_{[(2k+1)2^{-n-1},(2k+2)2^{-n-1}[}, \quad \forall t \in [0,1].$$

On vérifie que les fonctions h_0, h_n^k forment un système orthonormé de $L^2([0,1], \mathcal{B}([0,1]), \lambda)$ où λ désigne la mesure de Lebesgue. De plus ce système est total : toute fonction en escalier constante sur les intervalles de la forme $[i2^{-n}, (i+1)2^{-n}[$ (pour n fixé) est combinaison linéaire des fonctions h_0 et h_p^k pour $p < n$. On conclut que la famille

$$h_0, \; (h_n^k)_{n\geq 0, 0\leq k\leq 2^n-1}$$

forme une base orthonormée de $L^2([0,1], \mathcal{B}([0,1]), \lambda)$.

Notons $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$ le produit scalaire dans $L^2([0,1], \mathcal{B}([0,1]), \lambda)$. Alors, pour toute fonction $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$ on a

$$f = \langle f, h_0\rangle h_0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \langle f, h_n^k\rangle h_n^k.$$

D'autre part, nous disposons sur notre espace de probabilité $(\Omega, \mathcal{F}, P)$ d'une suite de v.a. $\mathcal{N}(0,1)$ indépendantes. Quitte à la renuméroter on peut écrire cette suite sous la forme

$$N_0, \; (N_n^k)_{n\geq 0, 0\leq k\leq 2^n-1}.$$

Il est immédiat de vérifier que cette famille constitue un système orthonormé dans $L^2(\Omega, \mathcal{F}, P)$. Il existe alors une (unique) isométrie, notée $\mathcal{B}$, de $L^2([0,1], \mathcal{B}([0,1]), \lambda)$ dans $L^2(\Omega, \mathcal{F}, P)$ telle que $\mathcal{B}(h_0) = N_0$ et $\mathcal{B}(h_n^k) = N_n^k$ pour tous $n \geq 0$, $0 \leq k \leq 2^n - 1$. Précisément,

$$\mathcal{B}(f) = \langle f, h_0 \rangle N_0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle N_n^k,$$

pour toute $f \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$ (la série converge dans $L^2(\Omega, \mathcal{F}, P)$). Remarquons que

$$E[\mathcal{B}(f)^2] = \|f\|_2^2$$

par la propriété d'isométrie, et que $E[\mathcal{B}(f)] = 0$ puisque les v.a. N_0, N_n^k sont toutes centrées. De plus le lemme suivant montrera que $\mathcal{B}(f)$ suit une loi gaussienne.

Lemme 14.2.2 *Soit (U_n) une suite de v.a. gaussiennes qui converge dans L^2 vers U. Alors U est aussi gaussienne.*

Preuve. Soit $m_n = E[U_n]$ et $\sigma_n^2 = \mathrm{var}(U_n)$. La convergence dans L^2 assure que $m_n \longrightarrow m = E[U]$ et $\sigma_n^2 \longrightarrow \sigma^2 = \mathrm{var}(U)$. Mais d'autre part, puisque la convergence dans L^2 entraîne la convergence en loi on a aussi pour tout $\xi \in \mathbb{R}$,

$$e^{im_n\xi - \sigma_n^2\xi^2/2} = E[e^{i\xi U_n}] \longrightarrow E[e^{i\xi U}]$$

ce qui montre que la fonction caractéristique de U s'écrit

$$E[e^{i\xi U}] = e^{im\xi - \sigma^2\xi^2/2}$$

et donc que U suit la loi $\mathcal{N}(m, \sigma^2)$. □

En écrivant

$$\mathcal{B}(f) = \lim_{m\to\infty} \Big(\langle f, h_0 \rangle N_0 + \sum_{n=0}^{m} \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle N_n^k \Big),$$

et en utilisant le fait qu'une combinaison linéaire de v.a. gaussiennes indépendantes est encore gaussienne, on déduit du lemme que $\mathcal{B}(f)$ suit la loi $\mathcal{N}(0, \|f\|_2^2)$. Remarquons aussi que, pour $f, f' \in L^2([0,1], \mathcal{B}([0,1]), \lambda)$,

$$\mathrm{cov}(\mathcal{B}(f), \mathcal{B}'(f)) = E[\mathcal{B}(f)\mathcal{B}(f')] = \langle f, f' \rangle$$

grâce à la propriété d'isométrie.

On pose alors, pour tout $t \in [0,1]$,

$$B_t = \mathcal{B}(\mathbf{1}_{[0,t]}).$$

En particulier, $B_0 = \mathcal{B}(\mathbf{1}_{\{0\}}) = \mathcal{B}(0) = 0$ p.s.

Vérifions d'abord que la famille $(B_t)_{t\in[0,1]}$ vérifie la propriété (P1), restreinte à l'intervalle de temps $[0,1]$. On se donne donc $0 = t_0 < t_1 < \cdots < t_p \leq 1$. Par linéarité, on a

$$B_{t_i} - B_{t_{i-1}} = \mathcal{B}(\mathbf{1}_{]t_{i-1},t_i]})$$

qui suit une loi $\mathcal{N}(0, t_i - t_{i-1})$. De plus, si $i \neq j$,

$$\mathrm{cov}(B_{t_i} - B_{t_{i-1}}, B_{t_j} - B_{t_{j-1}}) = E[(B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})] = \langle \mathbf{1}_{]t_{i-1},t_i]}, \mathbf{1}_{]t_{j-1},t_j]} \rangle = 0.$$

Or il est facile de vérifier que le vecteur $(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_p} - B_{t_{p-1}})$ est un vecteur gaussien : si $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$,

$$\sum_{j=1}^{p} \lambda_j (B_{t_j} - B_{t_{j-1}}) = \mathcal{B}\Big(\sum_{j=1}^{p} \lambda_j \mathbf{1}_{]t_{j-1},t_j]} \Big)$$

suit une loi gaussienne. D'après la Proposition 11.4.2, le fait que la matrice de covariance $(\mathrm{cov}(B_{t_i} - B_{t_{i-1}}, B_{t_j} - B_{t_{j-1}}))_{i,j=1,\ldots,p}$ soit diagonale entraîne l'indépendance des v.a. $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_p} - B_{t_{p-1}}$, ce qui achève la preuve de (P1).

Il reste à établir la propriété de continuité (P2). Pour l'instant, $B_t = \mathcal{B}(\mathbf{1}_{[0,t]})$ est défini comme un élément de $L^2(\Omega, \mathcal{F}, P)$, donc une classe d'équivalence de variables égales p.s. Pour que la vérification de (P2) ait un sens, il est nécessaire de spécifier un représentant dans cette classe d'équivalence, et cela pour chaque $t \in [0,1]$ (ce choix n'avait pas d'influence sur la validité ou non de (P1) mais il en a pour (P2)). A cette fin, nous allons étudier de plus près la série qui définit B_t. On commence par introduire les fonctions de Schauder

$$g_0(t) = \langle \mathbf{1}_{[0,t]}, h_0 \rangle = t$$
$$g_n^k(t) = \langle \mathbf{1}_{[0,t]}, h_n^k \rangle = \int_0^t h_n^k(s) ds.$$

Par construction, on a pour tout $t \in [0,1]$,

$$B_t = \mathcal{B}(\mathbf{1}_{[0,t]}) = tN_0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} g_n^k(t) N_n^k$$

où la série converge a priori dans $L^2(\Omega, \mathcal{F}, P)$ pour chaque $t \in [0,1]$ fixé. Nous allons montrer bien plus, à savoir que la série converge uniformément sur l'intervalle $[0,1]$, pour tout $\omega \in \Omega$, sauf peut-être pour ω appartenant à un ensemble $A \in \mathcal{F}$ de probabilité nulle. On définit alors $B_t(\omega)$ comme la somme de la série précédente si $\omega \in A^c$ et on prend $B_t(\omega) = 0$ pour tout $t \in [0,1]$ si $\omega \in A$ (puisque si une suite de v.a. converge p.s. et dans L^2 les limites p.s. et L^2 sont les mêmes, il est clair qu'on a ainsi simplement spécifié un choix dans la classe d'équivalence de v.a. égales p.s. à $\mathcal{B}(\mathbf{1}_{[0,t]})$, et on n'a rien changé à la validité de (P1)). On obtiendra la continuité des applications $t \to B_t(\omega)$ en observant qu'une limite uniforme de fonctions continues est continue.

On remarque d'abord que $0 \leq g_n^k \leq 2^{-n/2}$ et que pour n fixé les fonctions g_n^k, $0 \leq k \leq 2^n - 1$ sont à supports disjoints ($g_n^k(t) > 0$ seulement si $k2^{-n} < t < (k+1)2^{-n}$). Donc,

$$\sup_{t \in [0,1]} \Big| \sum_{k=0}^{2^n-1} g_n^k(t) N_n^k \Big| \leq 2^{-n/2} \sup_{0 \leq k \leq 2^n - 1} |N_n^k|.$$

Lemme 14.2.3 *Si N suit la loi $\mathcal{N}(0,1)$, on a pour tout $a \geq 1$,*

$$P(|N| \geq a) \leq e^{-a^2/2}.$$

Preuve. Il suffit d'écrire

$$P(|N| \geq a) = \frac{2}{\sqrt{2\pi}} \int_a^\infty dx\, e^{-x^2/2} \leq \frac{2}{\sqrt{2\pi}} \int_a^\infty dx\, \frac{x}{a}\, e^{-x^2/2} = \frac{2}{a\sqrt{2\pi}} e^{-a^2/2}.$$

□

Puisque les v.a. N_n^k sont toutes de loi $\mathcal{N}(0,1)$, on peut utiliser le lemme pour majorer

$$P\Big(\sup_{0\leq k\leq 2^n-1} |N_n^k| > 2^{n/4}\Big) \leq \sum_{k=0}^{2^n-1} P(|N_n^k| > 2^{n/4}) \leq 2^n \exp(-2^{\frac{n}{2}-1}).$$

En posant

$$A_n = \Big\{ \sup_{0\leq k\leq 2^n-1} |N_n^k| > 2^{n/4} \Big\}$$

on déduit du lemme de Borel-Cantelli et de l'estimation précédente que

$$P(\limsup A_n) = 0.$$

Donc si $A = \limsup A_n$ on a $P(A) = 0$ et d'autre part si $\omega \notin A$, alors pour tout n assez grand

$$\sup_{0\leq k\leq 2^n-1} |N_n^k| \leq 2^{n/4}$$

d'où

$$\sup_{t\in[0,1]} \Big| \sum_{k=0}^{2^n-1} g_n^k(t) N_n^k \Big| \leq 2^{-n/4}$$

ce qui assure que la série de la définition de B_t converge uniformément sur l'intervalle $[0,1]$. Cela termine la vérification de (P2). On peut aussi remarquer que cette construction donne $B_0(\omega) = 0$ pour tout $\omega \in \Omega$ et pas seulement p.s.

Il reste à s'affranchir de la restriction $t \in [0,1]$, et à généraliser le résultat en dimension d quelconque. Dans un premier temps on considère des familles $(B_t^{(1)})_{t\in[0,1]}, (B_t^{(2)})_{t\in[0,1]}$, etc. construites comme ci-dessus, en prenant à chaque fois une nouvelle suite de v.a. gaussiennes indépendantes, indépendante des suites précédentes. On pose ensuite

$$B_t = B_1^{(1)} + B_1^{(2)} + \cdots + B_1^{(k)} + B_{t-k}^{(k+1)} \qquad \text{si } t \in [k, k+1[.$$

On vérifie aisément que $(B_t)_{t\in\mathbb{R}_+}$ est un mouvement brownien en dimension un.

Pour passer à une dimension d quelconque, il suffit de se donner d mouvements browniens en dimension un indépendants, notés $(B_t^1)_{t\in\mathbb{R}_+}, \ldots, (B_t^d)_{t\in\mathbb{R}_+}$ et de poser

$$B_t = (B_t^1, B_t^2, \ldots, B_t^d)$$

pour tout $t \in \mathbb{R}_+$. Ceci achève la preuve du théorème. □

Si $x \in \mathbb{R}^d$, on appelle mouvement brownien issu de x tout processus $(B_t)_{t\in\mathbb{R}_+}$ tel que $(B_t - x)_{t\in\mathbb{R}_+}$ soit un mouvement brownien issu de 0.

14.3 La mesure de Wiener

Soit $C(\mathbb{R}_+, \mathbb{R}^d)$ l'espace des fonctions continues de $\mathbb{R}_+$ dans $\mathbb{R}^d$. On munit cet espace de la tribu $\mathcal{C}$ qui est la plus petite tribu rendant mesurables les applications coordonnées $\mathrm{w} \to \mathrm{w}(t)$ pour tout $t \in \mathbb{R}_+$.

Lemme 14.3.1 *La tribu $\mathcal{C}$ coïncide avec la tribu borélienne lorsque $C(\mathbb{R}_+, \mathbb{R}^d)$ est muni de la topologie de la convergence uniforme sur tout compact.*

Preuve. Soit $\mathcal{B}$ la tribu borélienne. L'inclusion $\mathcal{C} \subset \mathcal{B}$ découle de ce que les applications coordonnées sont continues donc mesurables pour la tribu boréliennes. Dans l'autre sens, rappelons qu'une distance sur $C(\mathbb{R}_+, \mathbb{R}^d)$ est fournie par

$$d(\mathrm{w}, \mathrm{w}') = \sum_{n=1}^{\infty} 2^{-n} \sup_{0 \le t \le n} (|\mathrm{w}(t) - \mathrm{w}'(t)| \wedge 1).$$

On sait que l'espace $C(\mathbb{R}_+, \mathbb{R}^d)$ est séparable et donc que tout ouvert est réunion dénombrable de boules. Il suffit alors de montrer que toute boule est dans la tribu $\mathcal{C}$, ou encore que pour $\mathrm{w}_0 \in C(\mathbb{R}_+, \mathbb{R}^d)$ fixé, l'application $\mathrm{w} \to d(\mathrm{w}_0, \mathrm{w})$ est $\mathcal{C}$-mesurable. Or en écrivant pour tout $n \ge 1$,

$$\sup_{t \in [0,n]} (|\mathrm{w}(t) - \mathrm{w}_0(t)| \wedge 1) = \sup_{t \in [0,n] \cap \mathbb{Q}} (|\mathrm{w}(t) - \mathrm{w}_0(t)| \wedge 1)$$

on obtient immédiatement cette propriété de mesurabilité. □

Définition 14.3.1 *Soit $(B_t)_{t \in \mathbb{R}_+}$ un mouvement brownien en dimension d (issu de 0), défini sur un espace de probabilité $(\Omega, \mathcal{F}, P)$. La mesure de Wiener en dimension d est la mesure de probabilité $\mathbb{P}_0$ sur $C(\mathbb{R}_+, \mathbb{R}^d)$ définie comme la mesure-image de $P(d\omega)$ par l'application*

$$\Phi : \quad \begin{array}{l} \omega \longrightarrow (B_t(\omega))_{t \in \mathbb{R}_+} \\ \Omega \longrightarrow C(\mathbb{R}_+, \mathbb{R}^d) \end{array}$$

Remarquons que l'application Φ est mesurable : comme cela a été observé dans le chapitre précédent dans un contexte un peu différent, il suffit de voir que la composée de Φ avec chacune des applications coordonnées $\mathrm{w} \to \mathrm{w}(t)$ est mesurable, ce qui est immédiat (cette composée donne les v.a. B_t).

La définition précédente n'a de sens que parce qu'elle ne dépend pas du choix du mouvement brownien B. Cela se voit de la manière suivante. Si $0 = t_0 < t_1 < \cdots < t_p$, on a pour tous $A_0, A_1, \ldots, A_p$ boréliens de $\mathbb{R}^d$,

$$\begin{aligned} &\mathbb{P}_0(\{\mathrm{w} \in C(\mathbb{R}_+, \mathbb{R}^d) : \mathrm{w}(t_0) \in A_0, \mathrm{w}(t_1) \in A_1, \ldots, \mathrm{w}(t_p) \in A_p\}) \\ &= P(B_{t_0} \in A_0, B_{t_1} \in A_1, \ldots, B_{t_p} \in A_p) \\ &= \mathbf{1}_{A_0}(0) \int_{A_1 \times \cdots \times A_p} dy_1 \ldots dy_p \, p_{t_1}(y_1) p_{t_2 - t_1}(y_2 - y_1) \cdots p_{t_p - t_{p-1}}(y_p - y_{p-1}), \end{aligned}$$

d'après la formule (14.2), qui est vraie pour n'importe quel mouvement brownien B (c'est juste une reformulation de (P1)). Or le lemme de classe monotone montre qu'une mesure de

probabilité sur $C(\mathbb{R}_+, \mathbb{R}^d)$ est caractérisée par ses valeurs sur les “cylindres”, c'est-à-dire les ensembles de la forme

$$\{\mathrm{w} \in C(\mathbb{R}_+, \mathbb{R}^d) : \mathrm{w}(t_0) \in A_0, \mathrm{w}(t_1) \in A_1, \ldots, \mathrm{w}(t_p) \in A_p\}.$$

Cela montre bien que $\mathbb{P}_0$ est déterminée de manière unique, indépendamment du choix du mouvement brownien B : autrement dit tous les mouvements browniens (issus de 0) ont la même loi, qui est la mesure de Wiener.

Remarque. En un certain sens, la mesure de Wiener joue sur l'espace $C(\mathbb{R}_+, \mathbb{R}^d)$ un rôle analogue à la mesure de Lebesgue sur $[0, 1]$.

Si $x \in \mathbb{R}^d$, on note aussi $\mathbb{P}_x(d\mathrm{w})$ la mesure-image de $\mathbb{P}_0(d\mathrm{w})$ par la translation $\mathrm{w} \to x + \mathrm{w}$ (c'est la loi du mouvement brownien issu de x).

Construction canonique du mouvement brownien. Elle consiste à prendre comme espace de probabilité $\Omega = C(\mathbb{R}_+, \mathbb{R}^d)$ muni de la tribu $\mathcal{C}$ et de la probabilité $\mathbb{P}_0$. On définit alors pour tout $t \geq 0$,

$$B_t(\mathrm{w}) = \mathrm{w}(t), \quad \forall \mathrm{w} \in \Omega.$$

La famille $(B_t)_{t \in \mathbb{R}_+}$, définie sur l'espace de probabilité $(\Omega, \mathcal{C}, \mathbb{P}_0)$, est un mouvement brownien issu de 0. La propriété (P2) est évidente. La propriété (P1) découle de la formule donnée ci-dessus pour

$$\mathbb{P}_0(\{\mathrm{w} \in C(\mathbb{R}_+, \mathbb{R}^d) : \mathrm{w}(t_0) \in A_0, \mathrm{w}(t_1) \in A_1, \ldots, \mathrm{w}(t_p) \in A_p\}).$$

De même, sous $\mathbb{P}_x$, $(B_t)_{t \in \mathbb{R}_+}$ est un mouvement brownien issu de x.

14.4 Premières propriétés du mouvement brownien

Dans ce paragraphe et le suivant, on considère un mouvement brownien B en dimension d, issu de 0. Pour tout $s \geq 0$ on note $\mathcal{F}_s$ la tribu engendrée par les v.a. $(B_r, 0 \leq r \leq s)$. On note aussi $\mathcal{F}_\infty$ la tribu engendrée par toutes les v.a. B_t, $t \in \mathbb{R}_+$.

Proposition 14.4.1 (i) *Si φ est une isométrie vectorielle de $\mathbb{R}^d$, $(\varphi(B_t))_{t \in \mathbb{R}_+}$ est aussi un mouvement brownien (en particulier $-B$ est un mouvement brownien);*

(ii) *pour tout $\gamma > 0$, le processus $B_t^\gamma = \frac{1}{\gamma} B_{\gamma^2 t}$ est aussi un mouvement brownien (invariance par changement d'échelle);*

(iii) *pour tout $s \geq 0$, le processus $B_t^{(s)} = B_{s+t} - B_s$ est un mouvement brownien indépendant de $\mathcal{F}_s$ (propriété de Markov simple).*

Preuve. (i) et (ii) sont très faciles. Pour l'indépendance dans (iii), on observe que pour tout choix de $t_1 < t_2 < \cdots < t_p$ et $r_1 < r_2 < \cdots < r_q \leq s$, la propriété (P1) entraîne que le vecteur

$$(B_{t_1}^{(s)}, \ldots, B_{t_p}^{(s)})$$

est indépendant de

$$(B_{r_1}, \dots, B_{r_q}).$$

En utilisant la Proposition 9.2.4, on en déduit aisément que la famille $(B_t^{(s)})_{t\in\mathbb{R}_+}$ est indépendante de $(B_r)_{0\le r\le s}$. $\square$

Théorème 14.4.2 (Loi du tout ou rien de Blumenthal) *Soit*

$$\mathcal{F}_{0+} = \bigcap_{s>0} \mathcal{F}_s.$$

La tribu $\mathcal{F}_{0+}$ est grossière, au sens où $\forall A \in \mathcal{F}_{0+}$, $P(A) = 0$ ou 1.

Preuve. Soit $A \in \mathcal{F}_{0+}$ et soient $t_1, \dots, t_p > 0$. Pour $\varepsilon > 0$ assez petit, la propriété de Markov simple (Proposition 14.4.1 (iii)) entraîne que $(B_{t_1} - B_\varepsilon, \dots, B_{t_p} - B_\varepsilon)$ est indépendant de $\mathcal{F}_\varepsilon$, donc a fortiori de $\mathcal{F}_{0+}$. En conséquence, pour toute fonction f continue bornée sur $(\mathbb{R}^d)^p$,

$$E[\mathbf{1}_A\, f(B_{t_1} - B_\varepsilon, \dots, B_{t_p} - B_\varepsilon)] = P(A)\, E[f(B_{t_1} - B_\varepsilon, \dots, B_{t_p} - B_\varepsilon)].$$

En faisant tendre ε vers 0 on trouve

$$E[\mathbf{1}_A\, f(B_{t_1}, \dots, B_{t_p})] = P(A)\, E[f(B_{t_1}, \dots, B_{t_p})],$$

et donc $(B_{t_1}, \dots, B_{t_p})$ est indépendant de $\mathcal{F}_{0+}$. Grâce à nouveau à la Proposition 9.2.4, il en découle que $\mathcal{F}_\infty$ est indépendante de $\mathcal{F}_{0+}$. En particulier $\mathcal{F}_{0+} \subset \mathcal{F}_\infty$ est indépendante d'elle-même, ce qui entraîne que $\mathcal{F}_{0+}$ est grossière. $\square$

Corollaire 14.4.3 *On suppose $d = 1$. Alors, p.s. pour tout $\varepsilon > 0$*

$$\sup_{0\le s\le\varepsilon} B_s > 0, \qquad \inf_{0\le s\le\varepsilon} B_s < 0.$$

Pour tout $a \in \mathbb{R}$, soit $T_a = \inf\{t \ge 0 : B_t = a\}$ $(\inf \varnothing = \infty)$. Alors,

$$p.s.,\ \forall a \in \mathbb{R}, \quad T_a < \infty.$$

En conséquence, p.s.,

$$\limsup_{t\to\infty} B_t = +\infty, \quad \liminf_{t\to\infty} B_t = -\infty.$$

Remarque. Il n'est pas a priori évident que la variable $\sup_{0\le s\le\varepsilon} B_s$ soit mesurable: il s'agit d'un supremum non dénombrable de fonctions mesurables. Cependant, parce que nous savons que les trajectoires de B sont continues, on peut se restreindre aux valeurs **rationnelles** de $s \in [0, \varepsilon]$ et on obtient un supremum dénombrable de variables aléatoires (ou alors on peut utiliser le Lemme 14.3.1).

Preuve. Soit (ε_p) une suite de réels strictement positifs décroissant vers 0, et soit

$$A = \bigcap_p \{\sup_{0\le s\le\varepsilon_p} B_s > 0\}.$$

Il est clair que l'événement A est $\mathcal{F}_{0+}$-mesurable. D'autre part,

$$P(A) = \lim_{p\to\infty} \downarrow P(\sup_{0\le s\le \varepsilon_p} B_s > 0),$$

et

$$P(\sup_{0\le s\le \varepsilon_p} B_s > 0) \ge P(B_{\varepsilon_p} > 0) = \frac{1}{2},$$

puisque B_{ε_p} suit la loi gaussienne $\mathcal{N}(0,\varepsilon_p)$ qui est symétrique. Cela montre que $P(A) \ge 1/2$. D'après le Théorème 14.4.2 on a $P(A) = 1$, d'où

$$\text{p.s. } \forall \varepsilon > 0, \quad \sup_{0\le s\le \varepsilon} B_s > 0.$$

L'assertion concernant $\inf_{0\le s\le\varepsilon} B_s$ est obtenue en remplaçant B par $-B$.

Ensuite, on écrit

$$1 = P(\sup_{0\le s\le 1} B_s > 0) = \lim_{\delta\downarrow 0} \uparrow P(\sup_{0\le s\le 1} B_s > \delta),$$

et on remarque en appliquant la propriété d'invariance d'échelle (Proposition 14.4.1 (ii)) avec $\gamma = \delta$ que

$$P(\sup_{0\le s\le 1} B_s > \delta) = P(\sup_{0\le s\le 1/\delta^2} B_s^\delta > 1) = P(\sup_{0\le s\le 1/\delta^2} B_s > 1)$$

(la dernière égalité est vraie parce que la loi du mouvement brownien est définie de manière unique : voir les remarques suivant la Définition 14.3.1). En faisant tendre δ vers 0, on trouve

$$P(\sup_{s\ge 0} B_s > 1) = 1.$$

A nouveau un argument de changement d'échelle montre que pour tout $A > 0$,

$$P(\sup_{s\ge 0} B_s > A) = 1$$

et en utilisant le changement $B \to -B$ on a aussi

$$P(\inf_{s\ge 0} B_s < -A) = 1.$$

Les dernières assertions du corollaire en découlent facilement: pour la dernière, on observe qu'une fonction continue $f : \mathbb{R}_+ \longrightarrow \mathbb{R}$ ne peut visiter tous les réels que si $\limsup_{t\to+\infty} f(t) = +\infty$, $\liminf_{t\to+\infty} f(t) = -\infty$. □

En utilisant la propriété de Markov simple, on déduit facilement du corollaire que p.s. la fonction $t \to B_t$ n'est monotone sur aucun intervalle non-trivial.

14.5 La propriété de Markov forte

Notre but est d'étendre la propriété de Markov simple (Proposition 14.4.1 (iii)) au cas où l'instant déterministe s est remplacé par un temps aléatoire T. Nous devons d'abord préciser la classe des temps aléatoires admissibles. On garde les notations $\mathcal{F}_t$ et $\mathcal{F}_\infty$ introduites ci-dessus.

Définition 14.5.1 *Une variable aléatoire T à valeurs dans $[0, \infty]$ est un temps d'arrêt si $\forall t \geq 0$, $\{T \leq t\} \in \mathcal{F}_t$.*

Remarque. Si T est un temps d'arrêt, pour tout $t \geq 0$,

$$\{T < t\} = \bigcup_{q \in \mathbb{Q} \cap [0,t[} \{T \leq q\}$$

est dans $\mathcal{F}_t$.
Exemple. En dimension $d = 1$, $T_a = \inf\{t \geq 0 : B_t = a\}$ est un temps d'arrêt. En effet

$$\{T_a \leq t\} = \{ \inf_{r \in \mathbb{Q} \cap [0,t]} |B_r - a| = 0\} \in \mathcal{F}_t.$$

Définition 14.5.2 *Soit T un temps d'arrêt. La tribu des événements antérieurs à T est*

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty ; \forall t \geq 0,\ A \cap \{T \leq t\} \in \mathcal{F}_t\}.$$

On vérifie facilement que les variables aléatoires T et $\mathbf{1}_{\{T<\infty\}} B_T$ sont $\mathcal{F}_T$-mesurables (pour la deuxième remarquer que

$$\mathbf{1}_{\{T<\infty\}} B_T = \lim_{n\to\infty} \sum_{i=0}^{\infty} 1_{\{i2^{-n} \leq T < (i+1)2^{-n}\}} B_{i2^{-n}},$$

puis que, pour tout $s \geq 0$, $B_s \mathbf{1}_{\{s \leq T\}}$ est $\mathcal{F}_T$ mesurable).

Théorème 14.5.1 (Propriété de Markov forte) *Soit T un t.a. tel que $P(T < \infty) > 0$. Alors, conditionnellement à $\{T < \infty\}$, le processus $B^{(T)}$ défini par*

$$B_t^{(T)} = B_{T+t} - B_T$$

est un mouvement brownien indépendant de $\mathcal{F}_T$.

Remarque. Pour être tout à fait précis, il faut aussi définir $B^{(T)}$ sur l'ensemble $\{T = \infty\}$, par exemple en posant $B_t(\omega) = 0$ pour tout $t \geq 0$ si $T(\omega) = \infty$ (ce choix n'a évidemment aucune influence sur le résultat ci-dessus).

Preuve. Supposons d'abord $T < \infty$ p.s. On va montrer que, pour $A \in \mathcal{F}_T$, $0 \leq t_1 < \cdots < t_p$ et F continue bornée de $(\mathbb{R}^d)^p$ dans $\mathbb{R}_+$, on a

$$E[\mathbf{1}_A \, F(B_{t_1}^{(T)}, \ldots, B_{t_p}^{(T)})] = P(A) \, E[F(B_{t_1}, \ldots, B_{t_p})]. \tag{14.3}$$

Cela suffit pour établir les différentes assertions du théorème : le cas $A = \Omega$ montre que $B^{(T)}$ est un mouvement brownien (remarquer que les applications $t \to B_t^{(T)}(\omega)$ sont continues) et d'autre part (14.3) entraîne que pour tout choix de $0 \leq t_1 < \cdots < t_p$, le vecteur $(B_{t_1}^{(T)}, \ldots, B_{t_p}^{(T)})$ est indépendant de $\mathcal{F}_T$, d'où il découle que $B^{(T)}$ est indépendant de $\mathcal{F}_T$.

Pour montrer (14.3), on observe d'abord que p.s.

$$\begin{aligned} &F(B_{t_1}^{(T)}, \ldots, B_{t_p}^{(T)}) \\ &= \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbf{1}_{\{(k-1)2^{-n} < T \leq k2^{-n}\}} F(B_{k2^{-n}+t_1} - B_{k2^{-n}}, \ldots, B_{k2^{-n}+t_p} - B_{k2^{-n}}), \end{aligned}$$

d'où par convergence dominée,

$$\begin{aligned} &E[\mathbf{1}_A\, F(B_{t_1}^{(T)}, \ldots, B_{t_p}^{(T)})] \\ &= \lim_{n\to\infty} \sum_{k=0}^{\infty} E[\mathbf{1}_A \mathbf{1}_{\{(k-1)2^{-n} < T \leq k2^{-n}\}} F(B_{k2^{-n}+t_1} - B_{k2^{-n}}, \ldots, B_{k2^{-n}+t_p} - B_{k2^{-n}})]. \end{aligned}$$

Pour $A \in \mathcal{F}_T$, l'événement $A \cap \{(k-1)2^{-n} < T \leq k2^{-n}\}$ est $\mathcal{F}_{k2^{-n}}$-mesurable. D'après la propriété de Markov simple (Proposition 14.4.1 (iii)), on a donc

$$\begin{aligned} &E[\mathbf{1}_{A\cap\{(k-1)2^{-n} < T \leq k2^{-n}\}} F(B_{k2^{-n}+t_1} - B_{k2^{-n}}, \ldots, B_{k2^{-n}+t_p} - B_{k2^{-n}})] \\ &= P(A \cap \{(k-1)2^{-n} < T \leq k2^{-n}\})\, E[F(B_{t_1}, \ldots, B_{t_p})], \end{aligned}$$

et il ne reste plus qu'à sommer sur k pour arriver au résultat souhaité.

Lorsque $P(T = \infty) > 0$, les mêmes arguments conduisent à

$$E[\mathbf{1}_{A\cap\{T<\infty\}}\, F(B_{t_1}^{(T)}, \ldots, B_{t_p}^{(T)})] = P(A \cap \{T < \infty\})\, E[F(B_{t_1}, \ldots, B_{t_p})]$$

et le résultat recherché en découle à nouveau. □

Une application importante de la propriété de Markov forte est le principe de réflexion illustré dans la preuve du théorème suivant.

Théorème 14.5.2 *On suppose $d = 1$. Pour tout $t > 0$, notons $S_t = \sup_{s\leq t} B_s$. Alors, si $a \geq 0$ et $b \leq a$, on a*

$$P(S_t \geq a,\, B_t \leq b) = P(B_t \geq 2a - b).$$

En particulier, S_t a même loi que $|B_t|$.

Preuve. On applique la propriété de Markov forte au temps d'arrêt

$$T_a = \inf\{t \geq 0, B_t = a\}.$$

On a déjà vu (Corollaire 14.4.3) que $T_a < \infty$ p.s. Ensuite,

$$P(S_t \geq a,\, B_t \leq b) = P(T_a \leq t,\, B_t \leq b) = P(T_a \leq t,\, B_{t-T_a}^{(T_a)} \leq b - a),$$

puisque $B^{(T_a)}_{t-T_a} = B_t - B_{T_a} = B_t - a$. Notons $B' = B^{(T_a)}$, de sorte que d'après le théorème 14.5.1, le processus B' est un mouvement brownien indépendant de $\mathcal{F}_{T_a}$ donc en particulier de T_a. Comme B' a même loi que $-B'$, le couple (T_a, B') a aussi même loi que $(T_a, -B')$. Notons $H = \{(s, \mathrm{w}) \in \mathbb{R}_+ \times C(\mathbb{R}_+, \mathbb{R});\ s \le t,\ \mathrm{w}(t-s) \le b - a\}$. La probabilité précédente vaut

$$\begin{aligned}
P((T_a, B') \in H] &= P[(T_a, -B') \in H) \\
&= P(T_a \le t,\ -B^{(T_a)}_{t-T_a} \le b - a) \\
&= P(T_a \le t,\ B_t \ge 2a - b) \\
&= P(B_t \ge 2a - b)
\end{aligned}$$

parce que l'événement $\{B_t \ge 2a - b\}$ est contenu dans $\{T_a \le t\}$.

Pour la deuxième assertion on observe que

$$P(S_t \ge a) = P(S_t \ge a, B_t \ge a) + P(S_t \ge a, B_t \le a) = 2P(B_t \ge a) = P(|B_t| \ge a),$$

d'où le résultat voulu. □

On déduit immédiatement du théorème précédent que la loi du couple (S_t, B_t) a pour densité

$$g(a,b) = \frac{2(2a-b)}{\sqrt{2\pi t^3}} \exp\left(-\frac{(2a-b)^2}{2t}\right) \mathbf{1}_{\{a>0, b<a\}}.$$

Corollaire 14.5.3 $(d = 1)$ *Pour tout $a > 0$, T_a a même loi que $\dfrac{a^2}{B_1^2}$ et a donc pour densité*

$$f(t) = \frac{a}{\sqrt{2\pi t^3}} \exp\Big(-\frac{a^2}{2t}\Big) \mathbf{1}_{\{t>0\}}.$$

Preuve. On écrit

$$\begin{aligned}
P(T_a \le t) &= P(S_t \ge a) \\
&= P(|B_t| \ge a) \qquad \text{(Théorème 14.5.2)} \\
&= P(B_t^2 \ge a^2) \\
&= P(tB_1^2 \ge a^2) \qquad (B_t \text{ a même loi que } \sqrt{t}B_1) \\
&= P(\frac{a^2}{B_1^2} \le t).
\end{aligned}$$

Ensuite, puisque B_1 suit une loi $\mathcal{N}(0,1)$ on calcule facilement la densité de a^2/B_1^2. □

Reformulation sur l'espace canonique.

En vue des applications qui suivent, il sera utile de reformuler la propriété de Markov sur l'espace canonique $\Omega = C(\mathbb{R}_+, \mathbb{R}^d)$. A partir de maintenant on se place donc sur cet espace, sur lequel on considère le processus $B_t(\mathrm{w}) = \mathrm{w}(t)$, et la filtration canonique $\mathcal{F}_t = \sigma(B_s, 0 \le s \le t)$. Rappelons que, pour tout $x \in \mathbb{R}^d$, $(B_t)_{t\ge0}$ est sous $\mathbb{P}_x$ un mouvement brownien issu de x.

On introduit aussi les opérateurs de translation. Pour tout $s \geq 0$, $\theta_s : \Omega \longrightarrow \Omega$ est défini par

$$(\theta_s \mathrm{w})(t) = \mathrm{w}(s+t) \ , \ \forall t \geq 0.$$

Alternativement, $B_t \circ \theta_s = B_{s+t}$.

Théorème 14.5.4 *Soit T un temps d'arrêt, et soient F et G deux fonctions mesurables positives sur Ω. On suppose que F est $\mathcal{F}_T$-mesurable. Alors, pour tout $x \in \mathbb{R}^d$,*

$$\mathbb{E}_x[\mathbf{1}_{\{T<\infty\}} F \cdot G \circ \theta_T] = \mathbb{E}_x[\mathbf{1}_{\{T<\infty\}} F \cdot \mathbb{E}_{B_T}[G]].$$

Remarque. Comparer cet énoncé avec le Théorème 13.3.5.

Preuve. On se ramène facilement au cas $x = 0$. Pour alléger l'écriture supposons aussi $\mathbb{P}_0(T < \infty) = 1$. Le point-clé est d'observer que si $T(\mathrm{w}) < \infty$,

$$(\theta_T \mathrm{w})(t) = \mathrm{w}(T+t) = \mathrm{w}(T) + (\mathrm{w}(T+t) - \mathrm{w}(T)) = B_T(\mathrm{w}) + B_t^{(T)}(\mathrm{w}).$$

Ensuite on écrit

$$\mathbb{E}_0[\mathbf{1}_{\{T<\infty\}} F \cdot G \circ \theta_T] = \mathbb{E}_0[\mathbf{1}_{\{T<\infty\}} F \cdot G(B_T + B_\cdot^{(T)})] = \mathbb{E}_0[\mathbf{1}_{\{T<\infty\}} F \, \mathbb{E}_0[G(B_T + B_\cdot^{(T)}) \,|\, \mathcal{F}_T]],$$

où $B_\cdot^{(T)}$ désigne la fonction continue $(B_t^{(T)})_{t\geq 0}$, vue comme v.a. à valeurs dans $C(\mathbb{R}_+, \mathbb{R}^d)$. D'une part B_T est $\mathcal{F}_T$-mesurable, d'autre part $B_\cdot^{(T)}$ est indépendant de $\mathcal{F}_T$ et de loi $\mathbb{P}_0$, d'après le Théorème 14.5.1. En utilisant le Théorème 11.3.4, on a

$$\mathbb{E}_0[G(B_T + B_\cdot^{(T)}) \,|\, \mathcal{F}_T] = \int \mathbb{P}_0(d\mathrm{w}) \, G(B_T + \mathrm{w}) = \mathbb{E}_{B_T}[G]$$

d'où le résultat voulu. □

14.6 Fonctions harmoniques et problème de Dirichlet

Nous avons introduit dans le Chapitre 7 la mesure de Lebesgue sur la sphère S^{d-1} notée ω_d. La *mesure de probabilité uniforme* sur la sphère S^{d-1} est la mesure de probabilité σ_d obtenue en normalisant ω_d. D'après le Chapitre 7, σ_d est donc reliée à la mesure de Lebesgue λ_d sur $\mathbb{R}^d$ par la formule explicite

$$\sigma_d(A) = \frac{\Gamma(\frac{d}{2}+1)}{\pi^{d/2}} \lambda_d(\{rx : 0 \leq r \leq 1, \, x \in A\}),$$

pour tout borélien A de S^{d-1}. Comme ω_d, la mesure σ_d est invariante sous l'action des isométries vectorielles. De plus, le Théorème 7.2.1 donne la formule d'intégration en coordonnées polaires : pour toute fonction borélienne $f : \mathbb{R}^d \longrightarrow \mathbb{R}_+$,

$$\int_{\mathbb{R}^d} f(x) \, dx = c_d \int_0^\infty \int_{S^{d-1}} f(rz) \, r^{d-1} \, dr \, \sigma_d(dz). \tag{14.4}$$

avec $c_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$.

Lemme 14.6.1 *La mesure σ_d est la seule mesure de probabilité sur la sphère S^{d-1} qui soit invariante par l'action des isométries vectorielles.*

Preuve. Soit μ une autre mesure de probabilité sur S^{d-1} invariante par l'action des isométries vectorielles. Alors, pour tout $\xi \in \mathbb{R}^d$ et toute isométrie vectorielle Φ,

$$\widehat{\mu}(\xi) = \int e^{i\xi\cdot x}\mu(dx) = \int e^{i\xi\cdot\Phi^{-1}(x)}\mu(dx) \int e^{i\Phi(\xi)\cdot x}\mu(dx) = \widehat{\mu}(\Phi(\xi)).$$

Il en découle que $\widehat{\mu}(\xi)$ ne dépend que de $|\xi|$, et donc il existe une fonction $f : \mathbb{R}_+ \longrightarrow \mathbb{C}$ telle que, pour tout $\xi \in \mathbb{R}^d$,

$$\widehat{\mu}(\xi) = f(|\xi|).$$

Le même argument montre qu'il existe une fonction $g : \mathbb{R}_+ \longrightarrow \mathbb{C}$ telle que

$$\widehat{\sigma}_d(\xi) = g(|\xi|).$$

Alors, pour tout $r \geq 0$,

$$\int_{S^{d-1}} \Big(\int_{S^{d-1}} e^{ir\xi\cdot x}\mu(dx)\Big)\sigma_d(d\xi) = \int_{S^{d-1}} f(r)\,\sigma_d(d\xi) = f(r)$$

et d'après le théorème de Fubini cela est aussi égal à

$$\int_{S^{d-1}} \Big(\int_{S^{d-1}} e^{irx\cdot\xi}\sigma_d(d\xi)\Big)\mu(dx) = \int_{S^{d-1}} g(r)\,\mu(dx) = g(r).$$

Donc $f = g$, d'où $\widehat{\mu} = \widehat{\sigma}_d$ et $\mu = \sigma_d$ grâce au Théorème 8.2.4. □

Si $x \in \mathbb{R}^d$ et $r > 0$ on note $B(x, r)$ la boule ouverte de centre x et de rayon r, et $\bar{B}(x, r)$ la boule fermée. La probabilité uniforme sur la sphère de centre x et de rayon r, notée $\sigma_{x,r}$ est par définition l'image de $\sigma_d(dy)$ par l'application $y \to x + ry$.

Rappelons que jusqu'à la fin du chapitre on considère le mouvement brownien défini sur l'espace canonique comme cela a été précisé à la fin de la partie précédente.

Proposition 14.6.2 *Soit $x \in \mathbb{R}^d$ et $r > 0$, et soit S le temps d'arrêt*

$$S = \inf\{t \geq 0 : |B_t - x| \geq r\}.$$

La loi de B_S sous $\mathbb{P}_x$ est la probabilité uniforme $\sigma_{x,r}$.

Preuve. Modulo une translation et un changement d'échelle, il suffit de traiter le cas $x = 0, r = 1$, dans lequel $\sigma_{x,r} = \sigma_d$. Les propriétés d'invariance du mouvement brownien montrent que la loi de B_S est alors invariante par l'action des isométries vectorielles. Grâce au Lemme 14.6.1, la loi de B_S doit être σ_d. □

Rappelons qu'un domaine D est un ouvert connexe de $\mathbb{R}^d$. Une fonction $h : D \longrightarrow \mathbb{R}$ est dite localement bornée si elle est bornée sur tout sous-ensemble compact de D.

Définition 14.6.1 *Soit D un domaine de $\mathbb{R}^d$. Une fonction mesurable localement bornée $h : D \longrightarrow \mathbb{R}$ est dite harmonique si, pour tous $x \in D$ et $r > 0$ tels que la boule $\bar{B}(x,r)$ soit contenue dans D, on a*

$$h(x) = \int h(y)\, \sigma_{x,r}(dy). \tag{14.5}$$

En d'autres mots, la valeur de h en x coïncide avec sa moyenne sur la sphère de centre x et de rayon r, pourvu que la boule fermée $\bar{B}(x,r)$ soit contenue dans D.

Problème de Dirichlet classique. Etant donné un domaine borné D et une fonction continue $g : \partial D \longrightarrow \mathbb{R}$, on veut trouver une fonction $h : D \longrightarrow \mathbb{R}$ telle que :

- $h_{|\partial D} = g$ au sens où, pour tout $y \in \partial D$,

$$g(y) = \lim_{x \to y, x \in D} h(x) \ ;$$

- h est harmonique sur D.

Le théorème suivant fournit un candidat à la solution du problème de Dirichlet.

Théorème 14.6.3 *Soit D un domaine borné, et soit g une fonction mesurable bornée sur ∂D. Notons*

$$T = \inf\{t \geq 0 : B_t \notin D\}.$$

Alors la fonction

$$h(x) = \mathbb{E}_x[g(B_T)], \qquad x \in D$$

est harmonique sur D.

Ce théorème est bien sûr analogue à un résultat de la fin du chapitre précédent concernant les relations entre chaînes de Markov et fonctions harmoniques discrètes.

Preuve. En écrivant

$$\{T \leq t\} = \left\{ \inf_{0 \leq s \leq t, s \in \mathbb{Q}} \operatorname{dist}(B_s, D^c) = 0 \right\}$$

on voit que T est un temps d'arrêt. Des propriétés du mouvement brownien en dimension un il découle aussi que $T < \infty$ $\mathbb{P}_x$ p.s. On a vu qu'alors B_T est une variable aléatoire (même $\mathcal{F}_T$-mesurable) et donc $\mathbb{E}_x[g(B_T)]$ est bien définie, et bornée par $\sup\{|g(y)|, y \in \partial D\}$.

Justifions maintenant le fait que h est mesurable. Rappelons la notation $\mathcal{C}$ pour la tribu introduite sur $C(\mathbb{R}_+, \mathbb{R}^d)$. Alors, pour tout $A \in \mathcal{C}$, l'application $x \to \mathbb{P}_x(A)$ est mesurable : cela est vrai pour les cylindres de la forme $A = \{\mathrm{w} : \mathrm{w}(t_1) \in A_1, \ldots, \mathrm{w}(t_p) \in A_p\}$, puisque dans ce cas on a une formule explicite, et il suffit ensuite d'utiliser un argument de classe monotone. Il en découle que pour toute fonction F mesurable bornée sur $C(\mathbb{R}_+, \mathbb{R}^d)$, l'application $x \to \mathbb{E}_x[F]$ est mesurable. On applique ceci à

$$F(\mathrm{w}) = \mathbf{1}_{\{T(\mathrm{w}) < \infty\}} g(B_T(\mathrm{w})) = \mathbf{1}_{\{T(\mathrm{w}) < \infty\}} g(\mathrm{w}(T(\mathrm{w})))$$

et on obtient ainsi que h est mesurable.

Fixons maintenant $x \in D$ et $r > 0$ tels que $\bar{B}(x,r) \subset D$. Posons

$$S = \inf\{t \geq 0 : B_t \notin B(x,r)\} = \inf\{t \geq 0 : |B_t - x| \geq r\}.$$

Il est clair que $S \leq T$, $\mathbb{P}_x$ p.s. (en fait $S(\mathrm{w}) \leq T(\mathrm{w})$ pour tout $\mathrm{w} \in \Omega = C(\mathbb{R}_+, \mathbb{R}^d)$). De plus,

$$B_T = B_T \circ \theta_S \ , \quad \mathbb{P}_x \text{ p.s.}$$

En effet c'est simplement dire que si $t \to \mathrm{w}(t)$ est une "trajectoire" issue du point x, le point de sortie de D pour cette trajectoire est le même que celui pour la même trajectoire dont on a "effacé" le début entre le point de départ et le point de sortie de la boule $B(x,r)$: cela est évident parce que $\bar{B}(x,r) \subset D$.

On peut donc utiliser la propriété de Markov forte sous la forme du Théorème 14.5.4 et obtenir

$$h(x) = \mathbb{E}_x[g(B_T)] = \mathbb{E}_x[g(B_T) \circ \theta_S] = \mathbb{E}_x[\mathbb{E}_{B_S}[g(B_T)]] = \mathbb{E}_x[h(B_S)] = \int h(y)\, \sigma_{x,r}(dy)$$

la dernière égalité étant la Proposition 14.6.2. Cela termine la preuve. □

Pour montrer que la fonction h du théorème précédent est solution du problème de Dirichlet (sous l'hypothèse supplémentaire de continuité de g), il faudrait aussi montrer que, pour tout $y \in \partial D$,

$$g(y) = \lim_{x \to y, x \in D} \mathbb{E}_x[g(B_T)].$$

Intuitivement, si $x \in D$ est proche de $y \in \partial D$, le mouvement brownien partant de x va sortir rapidement de D, donc le point de sortie B_T sera proche de x, et aussi de y, et la continuité de g assurera que $g(B_T)$ est proche de $g(y)$ ce qui conduira au résultat voulu. Avant de rendre précis ce raisonnement, ce qui exigera certaines hypothèses supplémentaires, nous commençons par traiter la question de l'unicité de la solution.

La proposition suivante montre que les fonctions harmoniques sont automatiquement très régulières.

Proposition 14.6.4 *Si h est harmonique sur D, h est de classe C^∞ sur D. De plus, si $x \in D$ et $r > 0$ sont tels que $\bar{B}(x,r) \subset D$, on a*

$$h(x) = \frac{1}{\lambda_d(B(x,r))} \int_{B(x,r)} h(y)\, dy. \tag{14.6}$$

Preuve. Soit $r_0 > 0$, et soit

$$D_0 = \{x \in D : \mathrm{dist}(x, D^c) > r_0\}.$$

Il suffit de montrer que h est de classe C^∞ sur D_0. Pour cela, considérons une fonction $\phi : \mathbb{R} \to \mathbb{R}_+$ de classe C^∞ à support compact contenu dans $]0, r_0[$, et non identiquement nulle. Alors, pour tout $x \in D_0$ et tout $r \in]0, r_0[$,

$$h(x) = \int \sigma_{x,r}(dz)\, h(z) = \int \sigma_d(dy)\, h(x + ry).$$

On multiplie les deux membres extrêmes de cette égalité par $r^{d-1}\phi(r)$ et on intègre par rapport à dr entre 0 et r_0. En utilisant la formule (14.4) on trouve que, pour une constante $c > 0$ dépendant seulement de ϕ, on a pour tout $x \in D_0$,

$$\begin{aligned} c\,h(x) &= c_d \int_0^{r_0} dr\, r^{d-1}\phi(r) \int \sigma_d(dy)\, h(x+ry) \\ &= \int_{B(0,r_0)} dz\, \phi(|z|) h(x+z) \\ &= \int_{B(x,r_0)} dz\, \phi(|z-x|) h(x) \\ &= \int_{\mathbb{R}^d} dz\, \phi(|z-x|) \widetilde{h}(x) \end{aligned}$$

où pour la dernière égalité on a noté $\widetilde{h}$ la fonction obtenue en prolongeant h par la valeur 0 sur D^c (le choix de cette valeur n'intervient pas puisque si $x \in D_0$ et $z \in D^c$ on a $\phi(|z-x|) = 0$).

On voit ainsi que sur D_0, h coïncide avec la convolution de la fonction $z \to \phi(|z|)$, qui est de classe C^∞ et à support compact, avec la fonction $\widetilde{h}$, qui est mesurable bornée. Nous avons remarqué à la fin du Chapitre 2, comme application du théorème de dérivation sous le signe intégrale, qu'une telle convolution est de classe C^∞.

Il reste à établir la deuxième assertion. En reprenant le calcul ci-dessus avec $\phi = \mathbf{1}_{[0,r_0[}$, on trouve pour $x \in D_0$,

$$h(x) = c' \int_{B(x,r_0)} dy\, h(y)$$

où la constante c' dépend seulement de ϕ, donc seulement de r_0. En prenant $h = 1$ (qui est harmonique), on voit que $c' = (\lambda_d(B(x,r_0)))^{-1}$ d'où le résultat annoncé. □

Corollaire 14.6.5 *Si une solution du problème de Dirichlet existe, elle est unique.*

Preuve. Soient h_1 et h_2 deux solutions, et soit $f = h_1 - h_2$. Supposons f non identiquement nulle. Quitte à échanger les rôles de h_1 et h_2 on peut supposer que f prend des valeurs strictement positives. La fonction obtenue en prolongeant f par la valeur 0 sur ∂D est continue sur $\bar{D}$, et doit donc atteindre son maximum M dans D (rappelons que D est supposé borné et donc $\bar{D}$ est compact). Soit x_0 un point de D tel que $f(x_0) = M$. D'après la proposition précédente on a pour tout $r < \mathrm{dist}(x_0, D^c)$,

$$f(x_0) = \frac{1}{\lambda_d(B(x_0,r))} \int_{B(x_0,r)} dy\, f(y),$$

soit

$$\int_{B(x_0,r)} dy\, (f(x_0) - f(y)) = 0.$$

Puisque $f(x_0) \geq f(y)$ pour tout $y \in D$, ceci n'est possible que si $f(x_0) = f(y)$, $\lambda_d(dy)$ p.p. sur $B(x_0,r)$. Comme f est continue (à nouveau grâce à la proposition précédente) on a donc $f(x_0) = f(y)$ pour tout $y \in B(x_0,r)$. On a ainsi montré que $\{x \in D : f(x) = M\}$ est ouvert.

Mais d'autre part cet ensemble est aussi un fermé de D, et puisque D est connexe, on a nécessairement $\{x \in D : f(x) = M\} = D$. Cela est absurde puisque $M > 0$ et f doit tendre vers 0 à la frontière de D. □

Définition 14.6.2 *On dit que D satisfait la condition de cône extérieur si, pour tout $y \in \partial D$, il existe $r > 0$ et un cône de révolution ouvert C de sommet y tels que $C \cap B(y,r) \subset D^c$.*

Théorème 14.6.6 *Supposons que D est un domaine borné satisfaisant la condition de cône extérieur, et soit g une fonction continue sur ∂D. Alors la fonction*

$$h(x) = \mathbb{E}_x[g(B_T)], \quad x \in D$$

est l'unique solution du problème de Dirichlet.

Preuve. Compte-tenu du Théorème 14.6.3 et du Corollaire 14.6.5, il suffit de vérifier que, pour tout $y \in \partial D$ fixé,

$$\lim_{x \to y, x \in D} h(x) = g(y). \tag{14.7}$$

Soit $\varepsilon > 0$. Grâce à la continuité de g, on peut choisir $\delta > 0$ tel que, si $z \in \partial D$ et $|z-y| < \delta$, on a

$$|g(z) - g(y)| < \frac{\varepsilon}{3}.$$

Soit ensuite $M > 0$ tel que $|g(z)| < M$ pour tout $z \in \partial D$. On a alors, pour tout $\eta > 0$,

$$\begin{aligned} |\mathbb{E}_x[g(B_T)] - g(y)| &\le \mathbb{E}_x[|g(B_T) - g(y)|\mathbf{1}_{\{T\le\eta\}}] + \mathbb{E}_x[|g(B_T) - g(y)|\mathbf{1}_{\{T>\eta\}}] \\ &\le \mathbb{E}_x[|g(B_T) - g(y)|\mathbf{1}_{\{T\le\eta\}}\mathbf{1}_{\{\sup_{t\le\eta}|B_t - x|\le\delta/2\}}] \\ &\qquad +2M\mathbb{P}_x\Big(\sup_{t\le\eta}|B_t - x| > \frac{\delta}{2}\Big) + 2M\,\mathbb{P}_x(T>\eta) \\ &= I + II + III. \end{aligned}$$

Nous allons majorer séparément les trois termes I, II, III.

Si $|x-y| < \frac{\delta}{2}$, on a sur l'événement $\{T \le \eta\} \cap \{\sup_{t\le\eta}|B_t - x| \le \delta/2\}$

$$|B_T - y| \le |B_T - x| + |x-y| < \delta$$

et le choix de δ assure que le terme I est majoré par $\varepsilon/3$.

En utilisant l'invariance par translation, on a

$$II = 2M\mathbb{P}_0\Big(\sup_{t\le\eta}|B_t| > \frac{\delta}{2}\Big)$$

et donc le terme II ne dépend pas de x. Clairement II tend vers 0 quand η tend vers 0 (c'est juste dire que $\sup_{t\le\eta}|B_t| \longrightarrow 0$ en probabilité sous $\mathbb{P}_0$, ce qui est vrai puisqu'il y a convergence p.s. par continuité). On peut donc choisir $\eta > 0$ assez petit de manière que $II < \varepsilon/3$.

Comme ε a été choisi de manière arbitraire, il reste pour établir (14.7) à montrer qu'on peut choisir $\alpha \in]0, \delta/2]$ suffisamment petit de manière que si $|x-y| < \alpha$, le terme $III = 2M\,\mathbb{P}_x(T > \eta)$ est aussi majoré par $\varepsilon/3$. Or cela est une conséquence du lemme suivant, qui complète donc la preuve du théorème.

Lemme 14.6.7 *Sous la condition de cône extérieur, on a pour tout $y \in \partial D$ et tout $\eta > 0$,*

$$\lim_{x \to y, x \in D} \mathbb{P}_x(T > \eta) = 0.$$

Remarque. Comme cela a été suggéré après la preuve du Théorème 14.6.3, le point-clé dans la vérification de la condition frontière (14.7) est de s'assurer que le mouvement brownien partant près de la frontière de D va sortir de D rapidement, avec une grande probabilité. C'est précisément ce que nous dit le lemme. La condition de cône extérieur n'est pas la meilleure possible pour cela, mais elle donne déjà des applications intéressantes, comme nous le verrons plus loin.

Preuve. Commençons par réécrire la condition de cône extérieur en $y \in \partial D$. Pour $u \in S^{d-1}$ et $\gamma > 0$, notons

$$C(u, \gamma) = \{z \in \mathbb{R}^d : z \cdot u > (1 - \gamma)|z|\}$$

le cône de révolution ouvert de sommet 0, de direction u et d'ouverture γ. Alors on peut choisir $r > 0$, $u \in S^{d-1}$ et $\gamma > 0$ tels que

$$y + (C(u, \gamma) \cap B(0, r)) \subset D^c.$$

Pour alléger l'écriture on note $C = C(u, \gamma) \cap B(0, r)$. Posons aussi

$$\widetilde{C} = \{z \in \mathbb{R}^d : z \cdot u > (1 - \frac{\gamma}{2})|z|\} \cap B(0, \frac{r}{2})$$

qui correspond à l'intersection avec $B(0, \frac{r}{2})$ d'un cône "un peu plus petit" que $C(u, r)$.

Il découle facilement de la loi du tout ou rien (Théorème 14.4.2) que, si $T_{\widetilde{C}} = \inf\{t \geq 0 : B_t \in \widetilde{C}\}$, on a

$$T_{\widetilde{C}} = 0 \ , \ \mathbb{P}_0 \text{ p.s..}$$

En effet, si (ε_n) est une suite décroissant strictement vers 0, l'événement $\limsup\{B_{\varepsilon_n} \in \widetilde{C}\}$ est dans la tribu $\mathcal{F}_{0+}$, et un argument analogue à la preuve du Corollaire 14.4.3 montre que cet événement est de probabilité strictement positive.

Pour $a \in]0, r/2[$, notons

$$\widetilde{C}_a = \widetilde{C} \cap B(0, a)^c.$$

Puisque les ensembles $\widetilde{C}_a$ croissent vers $\widetilde{C}$ quand $a \downarrow 0$, on a $T_{\widetilde{C}_a} \downarrow T_{\widetilde{C}} = 0$, $\mathbb{P}_0$ p.s., et donc pour tout $\beta > 0$ on peut fixer a assez petit tel que

$$\mathbb{P}_0(T_{\widetilde{C}_a} \leq \eta) > 1 - \beta.$$

En utilisant le fait que $y + C \subset D^c$, on a, avec des notations évidentes,

$$\mathbb{P}_x(T \leq \eta) \geq \mathbb{P}_x(T_{y+C} \leq \eta) = \mathbb{P}_0(T_{y-x+C} \leq \eta).$$

Or un raisonnement géométrique simple (faire un dessin!) montre que, dès que $|y - x|$ est assez petit, le cône translaté $y - x + C$ contient $\widetilde{C}_a$, et alors

$$\mathbb{P}_x(T \leq \eta) \geq \mathbb{P}_0(T_{\widetilde{C}_a} \leq \eta) > 1 - \beta$$

d'après le choix de a. Comme β était arbitraire on a terminé la preuve du lemme. □

Nous en venons maintenant à une autre caractérisation analytique des fonctions harmoniques, qui est souvent prise comme définition.

Proposition 14.6.8 *Soit h une fonction localement bornée sur le domaine D. Alors h est harmonique sur D si et seulement si h est de classe C^2 sur D et $\Delta h = 0$.*

Preuve. On suppose d'abord que h est harmonique. La Proposition 14.6.4 montre que h est de classe C^∞ sur D. Soit $x \in D$ et soit $r_0 > 0$ tel que la boule $\bar{B}(x, r_0)$ soit contenue dans D. Toujours d'après la Proposition 14.6.4, on a pour tout $r \in]0, r_0]$,

$$h(x) = \frac{1}{\lambda_d(B(x,r))} \int_{B(x,r)} h(y)\, dy. \tag{14.8}$$

D'autre part la formule de Taylor à l'ordre deux montre que, pour $y \in B(x, r)$,

$$h(y) = h(x) + \sum_{i=1}^{d} \frac{\partial h}{\partial y_i}(x)\,(y_i - x_i) + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 h}{\partial y_i \partial y_j}(x)\,(y_i - x_i)(y_j - x_j) + o(r^2)$$

où le reste $o(r^2)$ est uniforme quand y décrit $B(x, r)$. En intégrant cette égalité sur $B(x, r)$, et en utilisant les symétries évidentes, on trouve

$$\int_{B(x,r)} h(y)\, dy = \lambda_d(B(x,r))\, h(x) + \frac{1}{2} \sum_{i=1}^{d} \frac{\partial^2 h}{\partial y_i^2}(x) \int_{B(x,r)} (y_i - x_i)^2 dy + o(r^{d+2}).$$

Posons $C_1 = \int_{B(0,1)} y_1^2 dy > 0$. L'égalité précédente et (14.8) conduisent à

$$\frac{C_1}{2}\, \Delta h(x)\, r^{d+2} + o(r^{d+2}) = 0$$

ce qui n'est possible que si $\Delta h(x) = 0$.

Inversement supposons h de classe C^2 sur D et $\Delta h = 0$. Il suffit alors de montrer que si U est une boule ouverte telle que $\bar{U} \subset D$, h est harmonique sur U. D'après le Théorème 14.6.6, il existe une (unique) fonction $\widetilde{h}$ continue sur $\bar{U}$, harmonique dans U, et telle que $\widetilde{h}(x) = h(x)$ pour tout $x \in \partial U$. De plus, la première partie de la preuve montre que $\Delta \widetilde{h} = 0$ sur U. En appliquant le lemme suivant aux deux fonctions $h - \widetilde{h}$ et $\widetilde{h} - h$ (définies sur $\bar{U}$) on trouve que $h = \widetilde{h}$ sur $\bar{U}$, ce qui termine la preuve de la proposition. □

Lemme 14.6.9 (Principe du maximum) *Soit V un ouvert borné de $\mathbb{R}^d$, et soit u une fonction continue sur $\bar{V}$, de classe C^2 dans V et telle que $\Delta u \geq 0$ sur V. Alors,*

$$\sup_{x \in \bar{V}} u(x) = \sup_{x \in \partial V} u(x).$$

Preuve. Supposons d'abord qu'on a la propriété plus forte $\Delta u > 0$ sur D. On raisonne par l'absurde en supposant

$$\sup_{x \in \bar{V}} u(x) > \sup_{x \in \partial V} u(x).$$

Dans ce cas on peut trouver $x_0 \in V$ tel que

$$u(x_0) = \sup_{x \in V} u(x).$$

On a alors

$$\frac{\partial u}{\partial y_j}(x_0) = 0 \ , \ \forall j \in \{1, \dots, d\}$$

et de plus la formule de Taylor à l'ordre deux assure que la matrice symétrique

$$M_{x_0} = \Big(\frac{\partial^2 u}{\partial y_i \partial y_j}(x_0) \Big)_{i,j \in \{1,\dots,d\}}$$

est négative, au sens où la forme quadratique associée ne prend que des valeurs négatives ou nulle. En particulier les valeurs propres de M_{x_0} sont toutes négatives ou nulles et la trace de M_{x_0} l'est aussi. Mais ceci est une contradiction puisque la trace de M_{x_0} est $\Delta u(x_0) > 0$.

Si on fait l'hypothèse plus faible $\Delta u \geq 0$ sur D, il suffit de poser pour tout $\varepsilon > 0$, et tout $x \in \bar{V}$

$$u_\varepsilon(x) = u(x) + \varepsilon x_1^2,$$

de sorte que $\Delta u_\varepsilon = \Delta u + 2\varepsilon > 0$. La première partie de la preuve assure que

$$\sup_{x \in \bar{V}} u_\varepsilon(x) = \sup_{x \in \partial V} u_\varepsilon(x),$$

et il ne reste plus qu'à faire tendre ε vers 0. □

14.7 Fonctions harmoniques et mouvement brownien

Nous commençons par un résultat important qui fait le lien entre fonctions harmoniques, mouvement brownien et martingales. Nous devons d'abord introduire la notion de martingale à temps continu, qui est une généralisation directe des martingales à temps discret étudiées dans le Chapitre 12. Rappelons que nous nous sommes placés sur l'espace canonique du mouvement brownien, décrit à la fin de la partie 3, et que $\mathcal{F}_t$ désigne sur cet espace la tribu engendrée par $(B_s, s \leq t)$. Une famille $(M_t)_{t \geq 0}$, indexée par les réels positifs, de v.a. intégrables est une martingale si M_t est $\mathcal{F}_t$-mesurable, pour tout $t \geq 0$, et si la relation $E[M_t \,|\, \mathcal{F}_s] = M_s$ est vraie pour tous $0 \leq s \leq t$.

Si U est un ouvert de $\mathbb{R}^d$, on note $H_U = \inf\{t \geq 0 : B_t \notin U\}$.

Théorème 14.7.1 *Soit D un domaine de $\mathbb{R}^d$. Une fonction continue $h : D \longrightarrow \mathbb{R}$ est harmonique si et seulement si pour tout ouvert borné U tel que $\bar{U} \subset D$ et U satisfait la condition de cône extérieur, le processus*

$$(h(B_{t \wedge H_U}))_{t \geq 0}$$

est une martingale sous $\mathbb{P}_x$, pour tout $x \in U$.

De manière informelle les fonctions harmoniques sont celles qui composées avec le mouvement brownien donnent des martingales. La condition de cône extérieur dans l'énoncé qui précède est superflue mais intervient pour des raisons techniques dans notre démonstration.

Preuve. Supposons d'abord que h est harmonique, et soit U un ouvert satisfaisant les conditions de l'énoncé. On note $H = H_U$ pour alléger, et on fixe $x \in U$. Remarquons que les

v.a. $h(B_{t\wedge H})$ sont bornées $\mathbb{P}_x$ p.s. par $\sup\{|h(y)| : y \in \bar{U}\} < \infty$. Soient $s \leq t$. Observons que la v.a. $B_{s\wedge H}$ est $\mathcal{F}_{s\wedge H}$-mesurable donc aussi $\mathcal{F}_s$-mesurable. Pour obtenir l'égalité recherchée $E[h(B_{t\wedge H})\,|\,\mathcal{F}_s] = h(B_{s\wedge H})$, il suffit de montrer que, pour toute v.a. F $\mathcal{F}_s$-mesurable bornée, on a

$$\mathbb{E}_x[F\,h(B_{s\wedge H})] = \mathbb{E}_x[F\,h(B_{t\wedge H})].$$

Or on peut interpréter h comme la solution (unique) du problème de Dirichlet dans U dont la condition frontière est simplement la restriction de h à ∂U. Le Théorème 14.6.6 montre que, pour tout $y \in U$,

$$h(y) = \mathbb{E}_y[h(B_H)].$$

Il en découle que

$$\mathbb{E}_x[F\,\mathbf{1}_{\{s<H\}}h(B_{s\wedge H})] = \mathbb{E}_x[F\,\mathbf{1}_{\{s<H\}}h(B_s)] = \mathbb{E}_x[F\,\mathbf{1}_{\{s<H\}}\,\mathbb{E}_{B_s}[h(B_H)]].$$

Mais puisque $F\,\mathbf{1}_{\{s<H\}}$ est $\mathcal{F}_s$-mesurable (exercice), la propriété de Markov (sous la forme du Théorème 14.5.4, avec le temps d'arrêt constant s) montre que

$$\mathbb{E}_x[F\,\mathbf{1}_{\{s<H\}}\,\mathbb{E}_{B_s}[h(B_H)]] = \mathbb{E}_x[F\,\mathbf{1}_{\{s<H\}}h(B_H)].$$

On obtient ainsi

$$\mathbb{E}_x[F\,h(B_{s\wedge H})] = \mathbb{E}_x[F\,\mathbf{1}_{\{s<H\}}h(B_s)] + \mathbb{E}_x[F\,\mathbf{1}_{\{s\geq H\}}h(B_H)] = \mathbb{E}_x[F\,h(B_H)].$$

Evidemment le même argument montre que

$$\mathbb{E}_x[F\,h(B_{t\wedge H})] = \mathbb{E}_x[F\,h(B_H)] = \mathbb{E}_x[F\,h(B_{s\wedge H})]$$

ce qui était l'égalité recherchée.

Dans l'autre sens, c'est plus simple. Si on suppose que h vérifie la propriété de l'énoncé, on prend pour U une boule ouverte dont l'adhérence est contenue dans D. La propriété de martingale permet d'écrire si $x \in U$

$$h(x) = \mathbb{E}_x[h(B_{t\wedge H})\,|\,\mathcal{F}_0] = \mathbb{E}_x[h(B_{t\wedge H})].$$

En faisant tendre t vers ∞, on a $h(x) = \mathbb{E}_x[h(B_H)]$, et le Théorème 14.6.3 montre que h est harmonique sur U ce qui suffit pour conclure. □

A partir de maintenant, on suppose que $d \geq 2$ (remarquer qu'en dimension un les fonctions harmoniques sont les fonctions affines).

Proposition 14.7.2 *Soient $0 \leq a < b$ et soit $D_{a,b}$ le domaine*

$$D_{a,b} = B(0,b)\backslash\bar{B}(0,a).$$

Soit $f : D_{a,b} \longrightarrow \mathbb{R}$ une fonction radiale, au sens où $f(x)$ ne dépend que de $|x|$. Alors f est harmonique si et seulement s'il existe deux constantes $C, C' \in \mathbb{R}$ telles que

$$f(x) = \begin{cases} C + C'\log|x| & \text{si } d = 2, \\ C + C'|x|^{2-d} & \text{si } d \geq 3. \end{cases}$$

Preuve. Nous savons déjà que f doit être de classe C^∞. Soit $g :]a,b[\longrightarrow \mathbb{R}$ la fonction telle que $f(x) = g(|x|)$. L'expression du Laplacien pour une fonction radiale montre que

$$\Delta f(x) = g''(|x|) + \frac{d-1}{|x|}\, g'(|x|).$$

D'après la Proposition 14.6.8, f est harmonique si et seulement si g satisfait l'équation différentielle

$$g''(r) + \frac{d-1}{r}\, g'(r) = 0$$

qu'il suffit de résoudre pour obtenir la proposition. □

Dans les deux énoncés suivants on note $T_A = \inf\{t \geq 0 : B_t \in A\}$ pour tout fermé A de $\mathbb{R}^d$.

Proposition 14.7.3 *Soit $x \in \mathbb{R}^d \backslash \{0\}$, et soient $\varepsilon, R > 0$ avec $\varepsilon < |x| < R$. Alors,*

$$\mathbb{P}_x(T_{\bar{B}(0,\varepsilon)} < T_{B(0,R)^c}) = \begin{cases} \frac{\log R - \log |x|}{\log R - \log \varepsilon} & si\ d = 2, \\ \frac{|x|^{2-d} - R^{2-d}}{\varepsilon^{2-d} - R^{2-d}} & si\ d \geq 3. \end{cases} \tag{14.9}$$

Remarque. L'énoncé analogue en dimension un est, pour $a < x < b$,

$$\mathbb{P}_x(T_a < T_b) = \frac{b-x}{b-a}$$

et se démontre exactement de la même manière (exercice).

Preuve. Considérons le domaine $D = D_{\varepsilon,R}$, qui vérifie la condition de cône extérieur, et soit g la fonction continue sur ∂D définie par

$$\begin{cases} g(y) = 1 & \text{si } |y| = \varepsilon, \\ g(y) = 0 & \text{si } |y| = R. \end{cases}$$

Alors le Théorème 14.6.6 montre que

$$h(x) = \mathbb{P}_x(T_{\bar{B}(0,\varepsilon)} < T_{B(0,R)^c})\ , \quad \varepsilon < |x| < R$$

est la solution unique du problème de Dirichlet avec condition frontière g. Mais en utilisant la Proposition 14.7.2, on voit immédiatement que le terme de droite dans (14.9) est solution du même problème de Dirichlet. Cela donne l'égalité recherchée. □

On peut déduire de la proposition précédente des informations intéressantes sur le comportement presque sûr des fonctions $t \longrightarrow B_t$.

Corollaire 14.7.4 (i) *Si $d \geq 3$, pour tous $\varepsilon > 0$ et $x \in \mathbb{R}^d$ tels que $\varepsilon < |x|$,*

$$\mathbb{P}_x(T_{\bar{B}(0,\varepsilon)} < \infty) = (\frac{\varepsilon}{|x|})^{d-2}.$$

De plus, pour tout $x \in \mathbb{R}^d$,

$$\lim_{t\to\infty} |B_t| = \infty\ , \qquad \mathbb{P}_x\ p.s.$$

(ii) *Si $d = 2$, pour tous $\varepsilon > 0$ et $x \in \mathbb{R}^d$ tels que $\varepsilon < |x|$,*

$$\mathbb{P}_x(T_{\bar{B}(0,\varepsilon)} < \infty) = 1$$

mais

$$\mathbb{P}_x(T_{\{0\}} < \infty) = 0.$$

De plus, $\mathbb{P}_x$ p.s., pour tout ouvert U de $\mathbb{R}^2$, l'ensemble $\{t \geq 0 : B_t \in U\}$ est non borné.

Par analogie avec le cas des chaînes de Markov, on dit que le mouvement brownien est transitoire en dimension $d \geq 3$ et récurrent en dimension $d = 2$. Noter que cette propriété de récurrence dans le plan n'entraîne pas que tous les points soient visités : au contraire un point fixé, autre que le point de départ, n'est pas visité avec probabilité 1.

Preuve. (i) La première assertion est facile puisque

$$\mathbb{P}_x(T_{\bar{B}(0,\varepsilon)} < \infty) = \lim_{n\uparrow\infty} \mathbb{P}_x(T_{\bar{B}(0,\varepsilon)} < T_{B(0,n)^c})$$

et il suffit d'appliquer la formule (14.9).

Ensuite, on pose pour tout entier $n \geq 1$

$$T_{(n)} = T_{B(0,2^n)^c}.$$

En appliquant la propriété de Markov forte en $T_{(n)}$ et en utilisant à nouveau la formule (14.9), on trouve, si $|x| \leq 2^n$,

$$\mathbb{P}_x\Big(\inf_{t\geq T_{(n)}} |B_t| \leq n\Big) = \mathbb{E}_x\Big[\mathbb{P}_{B_{T_{(n)}}}(T_{\bar{B}(0,n)} < \infty)\Big] = (\frac{n}{2^n})^{d-2}.$$

Le lemme de Borel-Cantelli entraîne alors que $\mathbb{P}_x$ p.s., pour tout entier n assez grand,

$$\inf_{t\geq T_{(n)}} |B_t| > n$$

et donc la fonction $t \to |B_t|$ converge vers ∞ quand $t \to \infty$.

(ii) D'après la formule (14.9) on a

$$\mathbb{P}_x(T_{\bar{B}(0,\varepsilon)} < T_{B(0,R)^c}) = \frac{\log R - \log|x|}{\log R - \log\varepsilon}$$

dès que $\varepsilon < |x| < R$. En faisant tendre R vers ∞ dans cette formule on trouve

$$\mathbb{P}_x(T_{\bar{B}(0,\varepsilon)} < \infty) = 1.$$

En faisant tendre ε vers 0 dans la même formule on obtient

$$\mathbb{P}_x(T_{\{0\}} < T_{B(0,R)^c}) = 0.$$

Comme $T_{B(0,R)^c} \uparrow \infty$ quand $R \uparrow \infty$, cela entraîne

$$\mathbb{P}_x(T_{\{0\}} < \infty) = 0.$$

On a donc à la fois

$$\mathbb{P}_x \text{ p.s. } \forall \varepsilon > 0,\ T_{\bar{B}(0,\varepsilon)} < \infty$$

et

$$\mathbb{P}_x \text{ p.s. } \ 0 \notin \{B_t : t \geq 0\}.$$

Ces deux propriétés entraînent que $\mathbb{P}_x$ p.s. 0 est un point d'accumulation de la fonction $t \to B_t$ quand $t \to \infty$. Donc, pour tout ouvert U contenant 0, l'ensemble $\{t \geq 0 : B_t \in U\}$ est $\mathbb{P}_x$ p.s. non borné. Un argument de translation donne alors la dernière propriété du corollaire, en remarquant aussi qu'on peut se limiter à une famillle dénombrable de choix de U. □

Noyau de Poisson. Rappelons que nous nous plaçons en dimension $d \geq 2$. Le noyau de Poisson (de la boule unité) est la fonction définie sur $B(0,1) \times S^{d-1}$ par

$$K(x,y) = \frac{1-|x|^2}{|x-y|^d}\ , \quad x \in B(0,1),\ y \in S^{d-1}.$$

Lemme 14.7.5 *Pour tout $y \in S^{d-1}$ fixé, la fonction $x \to K(x,y)$ est harmonique sur $B(0,1)$.*

Preuve. Posons $K_y(x) = K(x,y)$ pour $x \in B(0,1)$. Un calcul direct montre que $\Delta K_y = 0$ sur $B(0,1)$, et il suffit d'appliquer la Proposition 14.6.8. □

Lemme 14.7.6 *Pour tout $x \in B(0,1)$,*

$$\int_{S^{d-1}} K(x,y)\, \sigma_d(dy) = 1.$$

Preuve. Pour tout $x \in B(0,1)$, posons

$$F(x) = \int_{S^{d-1}} K(x,y)\, \sigma_d(dy).$$

Alors, on déduit facilement du lemme précédent que F est harmonique sur $B(0,1)$: on peut appliquer le théorème de Fubini pour vérifier que F satisfait la propriété de moyenne (ou dériver sous le signe intégrale pour montrer que $\Delta F = 0$). Par ailleurs, en utilisant les propriétés d'invariance de σ_d et de K par les isométries vectorielles, on obtient que F est une fonction radiale. Sur la boule ouverte privée de l'origine $B(0,1)\backslash\{0\}$, F doit donc être de la forme donnée dans la Proposition 14.7.2. Mais puisque F est aussi continue en 0, la constante C' intervenant dans les formules de cette proposition doit être nulle. On a donc, pour tout $x \in B(0,1)$, $F(x) = F(0) = 1$. □

Théorème 14.7.7 *Soit g une fonction continue sur S^{d-1}. La solution du problème de Dirichlet dans $B(0,1)$ avec condition frontière g est donnée par*

$$h(x) = \int_{S^{d-1}} K(x,y)\, g(y)\, \sigma_d(dy)\ ,\ x \in B(0,1).$$

De plus, pour tout $x \in B(0,1)$ fixé, la fonction $y \to K(x,y)$ est la densité par rapport à la mesure $\sigma_d(dy)$ de la loi sous $\mathbb{P}_x$ du point de sortie du mouvement brownien hors de $B(0,1)$.

Preuve. Les mêmes arguments que dans la preuve du Lemme 14.7.6 montrent que h est harmonique dans $B(0,1)$. Pour vérifier la condition frontière, fixons $y_0 \in S^{d-1}$. Pour tout $\delta > 0$, la forme explicite du noyau de Poisson montre que si $x \in B(0,1)$ et $y \in S^{d-1}$ sont tels que $|x - y_0| < \delta/2$ et $|y - y_0| > \delta$ on a

$$K(x,y) \leq (\frac{2}{\delta})^d (1 - |x|^2).$$

Il découle de cette majoration que, pour tout $\delta > 0$,

$$\lim_{x \to y_0, x \in B(0,1)} \int_{\{|y-y_0|>\delta\}} K(x,y)\, \sigma(dy) = 0. \tag{14.10}$$

Ensuite, si $\varepsilon > 0$ est donné, on choisit $\delta > 0$ assez petit pour que $|g(y) - g(y_0)| \leq \varepsilon$ dès que $y \in S^{d-1}$ et $|y - y_0| \leq \delta$. Si $M = \sup\{|g(y)| : y \in S^{d-1}\}$, il vient

$$\begin{aligned} |h(x) - g(y_0)| &= \Big| \int_{S^{d-1}} K(x,y)\, (g(y) - g(y_0))\, \sigma_d(dy) \Big| \\ &\leq 2M \int_{\{|y-y_0|>\delta\}} K(x,y)\, \sigma(dy) + \varepsilon, \end{aligned}$$

en utilisant le Lemme 14.7.6 pour la première égalité, et ensuite le choix de δ. Grâce à (14.10), on obtient maintenant

$$\limsup_{x \to y_0, x \in B(0,1)} |h(x) - g(y_0)| \leq \varepsilon.$$

Comme ε était arbitraire, cela donne bien la condition frontière voulue.

Enfin, pour la dernière assertion, on utilise le Théorème 14.6.6 qui affirme que la solution du même problème de Dirichlet est aussi donnée par

$$h(x) = \mathbb{E}_x[g(B_T)],$$

où $T = \inf\{t \geq 0 : B_t \notin D\}$. En comparant les deux formules pour h on obtient précisément que la loi de B_T est la mesure $K(x,y)\sigma_d(dy)$. □

Quelques références

Partie I : Intégration.

Le livre classique de Rudin [7] est toujours une bonne référence. Le livre de Briane et Pagès [2] est très détaillé et assez complet.

[1] M.R. Adams, V. Guillemin. *Measure Theory and Probability.* Birkhäuser, 1996.

[2] M. Briane, G. Pagès. *Théorie de l'Intégration.* Vuibert, 2000.

[3] D.L. Cohn. *Measure Theory.* Birkhäuser, 1980.

[4] J.L. Doob. *Measure Theory.* Springer, 1994.

[5] R.M. Dudley. *Real Analysis and Probability.* Chapman and Hall, 1989.

[6] D. Revuz. *Mesure et Intégration.* Hermann, 1994.

[7] W. Rudin. *Real and Complex Analysis.* McGraw Hill, 1974.

[8] D.W. Stroock. *A Concise Introduction to the Theory of Integration.* Birkhäuser, 1994.

Partie II : Probabilités.

[9] et [18] sont des ouvrages en français dont le niveau correspond grosso-modo à celui du cours. [10] et [11] sont des classiques dont la lecture est toujours intéressante. [13] et [17] sont des livres plus récents écrits par des probabilistes de tout premier plan.

[9] P. Barbe, M. Ledoux. *Probabilité.* Belin, 1998.

[10] P. Billingsley. *Probability and Measure*, 3rd ed. Wiley, 1995.

[11] L. Breiman. *Probability.* Addison-Wesley, 1968.

[12] K.L. Chung. *A Course in Probability Theory.* Harcourt Brace and World, 1968.

[13] R. Durrett. *Probability and Examples*, 2nd ed. Duxbury Press, 1996.

[14] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. I.* Wiley. (*Un grand classique sur tout ce que vous pouvez faire en probabilités sans théorie de la mesure*)

[15] G. Grimmett, D. Stirzaker. *Probability and Random Processes.* Oxford Science Publications, 1992.

[16] J. Neveu. *Bases Mathématiques du Calcul des Probabilités.* Masson, 1064. (*Livre de référence sur les outils de théorie de la mesure qui interviennent en probabilités*)

[17] J. Pitman. *Probability.* Springer, 1993.

[18] D. Revuz. *Probabilités.* Hermann, 1997.

[19] D.W. Stroock. *Probability Theory: An Analytic View.* Cambridge U. Press 1993. (*Livre plus avancé autour des liens entre analyse et probabilités*)

Partie III : Processus aléatoires.

[20] J. Neveu Martingales à temps discret. Masson 1972

[21] D. Williams Probability with martingales. Cambridge University Press 1991

[22] C. Dellacherie, P.A. Meyer Probabilités et potentiels, Chapitres V à VIII. Théorie des martingales. Hermann 1980 (*traite aussi et surtout le cas des martingales à temps continu*)

[23] P. Baldi, L. Mazliak, P. Priouret Martingales et chaînes de Markov. Hermann 1998

[24] K.L. Chung Markov chains with stationary transition probabilities. Springer 1967

[25] R. Durrett Essentials of stochastic processes. Springer 1999 (*donne beaucoup d'exemples concrets de chaînes de Markov*)

[26] D.W. Stroock An introduction to Markov processes. Springer 2005 (*pour une lecture plus avancée sur chaînes et processus de Markov*).